

GenCore version 5.1.3
copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 1, 2002, 12:14:41 ; Search time 26.5 seconds

62.872 Million cell updates/sec

Title: US-09-529-691A-1
Perfect score: 92
Sequence: I GVKGDKGNPGWPGAP 15

Scoring table: BLOSUM62
Gapov 10.0 , Gapext 0.5

Searched: 747574 seqs, 11107396 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%, Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802:*

- 1: /SIDS1/gcdata/geneseq/geneseqp-emb1/AA1981.DAT:*
- 2: /SIDS1/gcdata/geneseq/geneseqp-emb1/AA1982.DAT:*
- 3: /SIDS1/gcdata/geneseq/geneseqp-emb1/AA1983.DAT:*
- 4: /SIDS1/gcdata/geneseq/geneseqp-emb1/AA1984.DAT:*
- 5: /SIDS1/gcdata/geneseq/geneseqp-emb1/AA1985.DAT:*
- 6: /SIDS1/gcdata/geneseq/geneseqp-emb1/AA1986.DAT:*
- 7: /SIDS1/gcdata/geneseq/geneseqp-emb1/AA1987.DAT:*
- 8: /SIDS1/gcdata/geneseq/geneseqp-emb1/AA1988.DAT:*
- 9: /SIDS1/gcdata/geneseq/geneseqp-emb1/AA1989.DAT:*
- 10: /SIDS1/gcdata/geneseq/geneseqp-emb1/AA1990.DAT:*
- 11: /SIDS1/gcdata/geneseq/geneseqp-emb1/AA1991.DAT:*
- 12: /SIDS1/gcdata/geneseq/geneseqp-emb1/AA1992.DAT:*
- 13: /SIDS1/gcdata/geneseq/geneseqp-emb1/AA1993.DAT:*
- 14: /SIDS1/gcdata/geneseq/geneseqp-emb1/AA1994.DAT:*
- 15: /SIDS1/gcdata/geneseq/geneseqp-emb1/AA1995.DAT:*
- 16: /SIDS1/gcdata/geneseq/geneseqp-emb1/AA1996.DAT:*
- 17: /SIDS1/gcdata/geneseq/geneseqp-emb1/AA1997.DAT:*
- 18: /SIDS1/gcdata/geneseq/geneseqp-emb1/AA1998.DAT:*
- 19: /SIDS1/gcdata/geneseq/geneseqp-emb1/AA1999.DAT:*
- 20: /SIDS1/gcdata/geneseq/geneseqp-emb1/AA2000.DAT:*
- 21: /SIDS1/gcdata/geneseq/geneseqp-emb1/AA2001.DAT:*
- 22: /SIDS1/gcdata/geneseq/geneseqp-emb1/AA2001.DAT:*

ALIGNMENTS

RESULT 1

ID AAR12973 standard; peptide: 15 AA.

XX AC AAR12973;

XX DT 24-SEP-1991 (first entry)

XX DE IV-H1 based on type IV collagen alpha 1 chain triple helical region.

XX KW malignant cell growth; melanoma cell motility; cellular adhesion.

XX OS Synthetic.

XX PN W09108755-A.

XX PD 27-JUN-1991.

XX PR 06-DEC-1990; 90WO-US07162.

XX PR 14-DEC-1989; 89US-0450419.

SUMMARIES

12	92	100.0	39	19	AAW46043	Collagen sequence
13	92	100.0	43	18	AAW13024	Collagen alpha1(IV)
14	92	100.0	1669	22	AAM40863	Human polypeptide
15	92	100.0	1672	22	AAM3977	Lung cancer associ
16	86	93.5	406	21	AAB58169	Human alpha 5 (IV)
17	69	75.0	772	18	AAW23873	Human type IV coll
18	69	75.0	1685	22	ABG04839	Novel human diagno
19	69	75.0	1693	22	ABG15619	Novel human diagno
20	69	69.6	430	21	AAB58180	Lung cancer associ
21	64	69.6	224	22	ABB68977	Drosophila melanog
22	61	66.3	429	22	AAB16066	Human novel secrete
23	61	66.3	532	19	AAW40114	Human alpha-1(IV)
24	61	66.3	532	19	AAE02333	Human EST encoded
25	61	66.3	1690	22	AAB2534	Collagen alpha 1 (
26	60	65.2	1078	16	AAR71704	Type III procolla
27	60	65.2	1078	21	AAY96125	Collagen type III
28	60	65.2	1196	13	AAR28916	Collagen type III
29	60	65.2	1466	22	ABB64070	Bovine alpha1(III)
30	60	65.2	1466	22	ABB50291	Bovine alpha1(III)
31	60	65.2	1466	22	AAE02534	Porcine alpha1(III)
32	60	65.2	1466	22	AAE02537	Novel human diagno
33	60	65.2	1469	22	ABG1591	Drosophila melanog
34	60	65.2	1940	22	ABB64070	Partial Sequence O
35	59	64.1	471	16	AAR79163	Bovine type IV col
36	59	64.1	471	20	AAY4171	Bovine alpha1(IV)
37	59	64.1	471	21	AAY56783	Bovine alpha1(IV)
38	59	64.1	471	22	AAB09483	Bovine alpha1(IV)
39	59	64.1	546	22	ABG15617	Novel human diagno
40	59	64.1	547	22	ABG04840	Novel human diagno
41	59	64.1	1078	22	AAM40102	Human polypeptide
42	58	63.0	654	22	AGG63322	Amino acid sequenc
43	58	63.0	666	22	AGG63343	Amino acid sequenc
44	58	63.0	1288	20	AAW92296	Mouse alpha-1 (XVI)
45	58	63.0	1301	20	AAW92296	Human alpha-1 (XVI)

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Page 2

This polypeptide (designated IV-H1) corresponds to type IV collagen residues 1263-1277 from the major triple helical region of the α 1 chain type of IV collagen. It, or a polypeptide comprising this sequence, can be used to promote cellular attachment to substrata or to inhibit the metastasis and invasion of tumour cells.

Db 1 GVKGDDGGNPQWPQGP 15

Query Match 100.0%; Score 92; DB 12; Length 15;
 Best Local Similarity 100.0%; Pred. No. 5, 2e-07;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0

AC AAR89922;
XX DT 20-DEC-1995 (first entry)
XX

QY	1	GVKGDKGNPGPGAP	15
Db	1	GVKGDKGNPGPGAP	15

XX Cancer metastasis; adhesive peptide; core sequence; dextran; cancer;
KW water soluble polysaccharide; metastasis; wound; immunogenicity.
XX

AAR83639
ID AAR83639 standard; peptide; 15 AA.
XX

PN JPO70899999-A.
XX
PD 04 - APR - 1995.

DE Collagen IV alpha-2 chain NC1 domain peptide IV-H1 (1263-1277).
XX 11 GRK 1990 (MIST ENTRY)

XX
PR 17-SEP-1993; 93JP-0254779.
XX

XX KW proliferation; suppression; epithelial; glial; epiretinal; eye; membranes; contraction; migration; vitreoretinopathy; scarring; vitreoretinal; glaucoma.

DR
XX
PT
PT
WPI; 1995-167234/22.
Cancer metastasis inhibitory peptide derivs. - useful for inhibition of cancer metastasis, healing of wounds and regulation of

PN
XX
PD
WO9522979-A1
31-AUG-1995.

CC
PS
XX disclosure; page 3; opp; Japanese.

XX	28-FEB-1995;	93W0-US024/6
PR	27-FEB-1995;	95US-0203458,
PR	28-FEB-1994;	94US-0203458,
XX		

CC CC CC CC CC CC
 WITH A CORE SEQUENCE SELECTED FROM RGD (Arginine-Glycine-Aspartic acid) (AARRGGRG), or other sequence (ARRRGGRRG), linked to a water soluble polysaccharide, preferably a water soluble dextran, at the C-terminus. The peptides are useful in inhibiting cancer metastasis, healing wounds and the regulation of immunogenicity.

XX
pi
Agarwal A,
Balles M,
Furcitt LT,
Gregerson DS,
Murali S;
Skubitz AP;
Wright MM,

Query Match Best Local Similarity 100.0%; Score 92; DB 16; Length 15;

PT
PT
PT
PT
treating proliferation
glaucoma - using pep
collagen

QY 1 GYKGDKGNPGWPOAP 15
Dp 1 GYKGDKGNPGWCGAP 15

xx

RESULT 4
ANW01130

Sequence 15 AA:
SQ CC CC CC CC CC CC CC CC CC
fragment of the fibronectin A chain (AAR83679-94), C-terminal G domain of the laminin A chain (AAR83642/7) or NC1 domain of the alpha-2 chain of type IV collagen (AAR83393-41) suppress fibroblast, epithelial and glial cell proliferation, contraction of epiretinal membranes and cell migration within the eye. It may be used in the treatment of proliferative vitreoretinopathy, vitreoretinal scarring and glaucoma.

ID RAW01139 standard; peptide; 15 AA.
XX
AC AAW01139;
XX
DT 18-DEC-1996 (first entry)
XX
DE Peptide 10 for glia cell removal derived from collagen.
XX
KW glia; neuron; analysis; behaviour; selective; removal.

XX	Synthetic.
OS	
XX	
FH	
FT	
Key	location/Qualifier
Modified-site	15 note= "PRO-NH2"

QY 1 GVKGDKGNPGWPGAP 15
 |||||
 ID AAY04481 standard; peptide; 15 AA.
 Db 1 GVKGDKGNPGWPGAP 15

RESULT 7

AAY04481
 ID AAY04481 standard; peptide; 15 AA.
 XX
 AC
 XX
 DT 05-JUL-1999 (first entry)
 DE Human type IV collagen alpha 1 chain D form fragment D-IVHL.
 XX
 KW Human; type IV collagen alpha 1 chain; D-IVHL; inhibition; tumour;
 cell adhesion; major triple helical domain; collagenous region.
 XX
 OS Homo sapiens.
 OS Synthetic.

FH Location/Qualifiers
 FT Misc-difference 1..15 /note= "D-form residues"
 FT PT
 PN WO9920300-A1.
 XX
 PD 29-APR-1999.
 XX
 PF 22-OCT-1998; 98WO-US22405.
 PR 22-OCT-1997; 97US-0062716.
 PR 22-OCT-1997; 97US-0062617.
 PA (MINU) UNIV MINNESOTA.
 XX
 PI Fields GB, McCarthy JB;
 XX
 DR WPI; 1999-302644/25.
 XX
 PT New polypeptide useful for inhibition of tumour cell adhesion to type
 IV collagen
 XX
 PS Claim 4; Page 18; 31pp; English.
 XX
 CC The present sequence represents an all D-form fragment of the continuous
 collagenous region of the major triple helical domain of the alpha 1
 chain of human type IV collagen, designated D-IVHL. The peptide D-IVHL
 can be used to inhibit tumour cell binding to type IV collagen, inhibit
 tumour cell invasion of basement membrane and inhibit tumour cell
 metastasis in vivo. The peptide can be used to treat melanoma cell
 metastasis, and spontaneous Lewis lung tumour metastasis. The peptide is
 also useful for targeting cytotoxic agents and drugs to tumour cells.
 XX
 SQ Sequence 15 AA;

Query Match 100.0%; Score 92; DB 21; Length 15;
 Best Local Similarity 100.0%; Pred. No. 5.2e-07;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVKGDKGNPGWPGAP 15
 |||||
 ID 1 GVKGDKGNPGWPGAP 15
 Db 1 GVKGDKGNPGWPGAP 15

RESULT 8

Query Match 100.0%; Score 92; DB 20; Length 15;
 Best Local Similarity 100.0%; Pred. No. 5.2e-07;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

FH Location/Qualifiers
 FT Modified-site 3 /label= 4HYP
 FT Modified-site 6 /label= 4HYP
 FT Modified-site 9 /label= 4HYP
 FT Modified-site
 ID AAY78574 standard; peptide; 15 AA.
 XX
 AC
 XX
 DT 05-MAY-2000 (first entry)
 XX

DE Type IV collagen derived peptide IV-HI.
 XX
 KW Type IV collagen; intraocular fibroblast proliferation inhibitor;
 KW proliferative retinopathy; fibroblast migration; scarring; glaucoma.
 OS Unidentified.
 XX
 PN US6013628-A.
 XX
 PD 11-JAN-2000.
 XX
 PF 27-FEB-1995; 95US-0394748.
 PR 28-FEB-1994; 94US-0203458.
 XX
 PA (MINU) UNIV MINNESOTA.
 PI Gregerson DS, Agarwal A, Wright MM, Murali S, Skubitz APN;
 Furcht LT, Ballies M;
 DR WPI; 2000-159882/14.
 XX
 PT Treatment of glaucoma by ocular administration of a type IV collagen
 peptide that inhibits fibroblast proliferation -
 PT Example 1; Column 15; 44pp; English.
 XX
 CC This sequence represents a type IV collagen peptide. Ocular
 administration of the peptide inhibits fibroblast proliferation and can
 be used in the treatment of glaucoma. The peptide can be used to treat
 proliferative vitreoretinopathy by ocular administration or the peptides
 of the invention, which suppress fibroblast-mediated collagen gel
 contraction. The peptides can be used for treating glaucoma in mammals,
 especially humans, by inhibiting intraocular fibroblast proliferation,
 CC fibroblast migration and scarring.
 XX
 SQ Sequence 15 AA;

Query Match 100.0%; Score 92; DB 21; Length 15;
 Best Local Similarity 100.0%; Pred. No. 5.2e-07;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVKGDKGNPGWPGAP 15
 |||||
 ID AAW46044 standard; peptide; 24 AA.
 XX
 AC AAW46044;
 XX
 DT 23-DEC-1998 (first entry)
 DE Collagen sequence synthesised as triple-helical peptide.
 XX
 KW Triple helix; collagen structure; homo-trimer; hetero-trimer.
 OS Synthetic.
 XX
 FH Location/Qualifiers
 FT Modified-site 3 /label= 4HYP
 FT Modified-site 6 /label= 4HYP
 FT Modified-site 9 /label= 4HYP
 FT Modified-site
 ID AAY78574 standard; peptide; 15 AA.
 XX
 AC
 XX
 DT 10-MAR-1998.
 XX

PF 03-JUL-1996; 96US-0675140.
 XX
 PR 30-JUN-1993; 93US-0085633.
 PR 27-SEP-1995; 95US-0534342.
 PR 03-JUL-1996; 96US-0675140.
 XX
 PA (MINN) UNIV MINNESOTA.
 XX
 PT Fields GB;
 XX
 DR WPI; 1998-192815/17.
 XX
 PT Supported polypeptide(s) useful as intermediates - for
 triple-helical polypeptide(s) having collagen structures
 XX
 PS disclosure; Column 3; 23pp; English.
 XX
 CC The invention relates to supported polypeptides of formula:
 A-B-B-(J){^uL-R; where A = an N alpha-amino protecting group
 CC removable under non-acidic conditions; B = an amino acid having a single
 CC side-chain amino group protected with a group removable under non-strong
 CC acidic conditions; J = an amino acid that is capable of acting as a
 chromophore and is protected with a group capable of withstanding the
 conditions under which the protecting groups of A and B are removed; u =
 CC an amino acid; u = 0 or 1; J = 1 or more; L = a linker capable of being
 labelled using a non-strong acid mechanism, and R = a support material.
 CC The supported polypeptides are intermediates for triple-helical
 CC polypeptides. The triple-helical polypeptides are useful for studying the
 structure and biological activity of specific collagen sequences in homo-
 CC and hetero-trimeric forms. The present sequence represents a collagen
 CC sequence used in the synthesis of triple-helical peptides.
 XX
 SQ Sequence 24 AA;

Query Match	100.0%	Score 92;	DB 19;	Length 24;
Best Local Similarity	100.0%	Pred. No.	8.5e-07;	
Matches	15;	Conservative	0;	Mismatches 0;
Qy	1	GVKGDKGNPGWPGAP 15		
Db	10	GVKGDKGNPGWPGAP 24		

RESULT 10
 AAW13025
 ID AAW13025 standard; peptide; 27 AA.
 XX
 AC AAW13025;
 XX
 DT 30-APR-1997 (first entry)
 XX
 DE Collagen alpha(IIV) chain containing triple helical polypeptide 4.
 XX
 KW Triple helical polypeptide; collagen; cell; adhesion; migration;
 KW promotion; alpha(IIV) chain; solid phase synthesis; study;
 KW O-glycosidic bond; deuterium label; structure; biological activity;
 KW homotrimeric; heterotrimeric; nuclear magnetic resonance; NMR.
 XX
 OS Synthetic.
 XX
 PH Key
 PT Modified-site 3
 PT /label= Hyp
 PT Modified-site 6
 PT /label= Hyp
 PT Modified-site 9
 PT /label= Hyp
 PT Peptide 10. 24
 PT /note= "collagen alpha(IIV) chain residues
 PT 1263-1277"
 PT Modified-site 25
 PT /note= "linked via peptide bond to Pro24 of peptide comprising residues 1-24 of present

PT Modified-site 26 sequence"
 PT /note= "linked via peptide bond to Pro24 of peptide comprising residues 1-24 of present sequence."
 PT sequence".
 PT
 PN US5576419-A.
 XX
 PD 19-NOV-1996.
 XX
 PP 30-JUN-1993; 93US-0085633.
 PR 30-JUN-1993; 93US-0085633.
 PR 27-SEP-1995; 95US-0534342.
 XX
 PA (MINN) UNIV MINNESOTA.
 XX
 PI Fields GB;
 XX
 DR WPI; 1997-011309/01.
 XX
 PT Solid phase synthesis of triple-helical branched polypeptide -
 PT under O-glycosidic bond and deuterium label retaining conditions,
 PT which may contain collagen cell adhesion sequences, useful for
 PT studying structure and biological activity of collagen
 XX
 PS Claim 52; Column 27; 25pp; English.
 XX
 CC The present sequence is that of triple helical polypeptide 4
 CC (THP-4), which includes a collagen cell adhesion sequence, known to
 CC promote adhesion and migration of various cells (residues 1263-1277
 CC of the collagen alpha(IIV) chain). In its triple helix inducing
 CC region, THP-4 was prepared by solid phase synthesis under
 CC O-glycosidic bond and deuterium label retaining conditions, without
 CC strong acidolytic steps which affect O-glycosylation and isotopic
 CC labelling. THP can be used to study the structure and biological
 CC activity of particular collagens as homotrimeric or heterotrimeric
 CC forms, with the advantage that different adhesion sequences can be
 CC studied with a minimum of Gly-Pro-Hyp triplets, avoiding
 CC overlapping NMR resonances.
 XX
 SQ Sequence 27 AA;

Query Match	100.0%	Score 92;	DB 18;	Length 27;
Best Local Similarity	100.0%	Pred. No.	9.6e-07;	
Matches	15;	Conservative	0;	Mismatches 0;
Qy	1	GVKGDKGNPGWPGAP 15		
Db	10	GVKGDKGNPGWPGAP 24		

RESULT 11
 AAW46046
 ID AAW46046 standard; peptide; 39 AA.
 XX
 AC AAW46046;
 XX
 DT 23-DEC-1998 (first entry)
 XX
 DE Peptide intermediate used in synthesis of triple-helical peptide.
 XX
 KW Triple helix; collagen structure; homo-trimer; hetero-trimer.
 XX
 OS Synthetic.
 XX
 PH Key
 PT Modified-site 3
 PT /label= 4Hyp
 PT Modified-site 6
 PT /label= 4Hyp
 PT Modified-site 9
 PT /label= 4Hyp

AAW13024	DE	Collagen alpha1(IV) chain containing triple helical polypeptide 2.	CC	strong acidolysis steps which affect O-glycosylation and isotopic
AAW13024 standard; peptide; 43 AA.	KW	labeling. This can be used to study the structure and biological	CC	activity of particular collagens as homotrimeric or heterotrimeric
AC	KW	forms, with the advantage that different adhesion sequences can be	CC	studied with a minimum of Gly-Pro-Hyp triplets, avoiding
XX	KW	overlapping NMR resonances.	CC	
DT	30-APR-1997 (first entry)		XX	
XX	OS	synthetic.	SQ	Sequence 43 AA;
XX	FH		Query Match 100.0%; Score 92; DB 18; Length 43;	
FT	Key	Location/Qualifiers	Best Local Similarity 100.0%; Prod. No. 1.6e-06;	
FT	Modified-site	3	Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
FT	Modified-site	/label= HYP	Db 25 GVKGDKGNPGWPAG 39	
FT	Modified-site	6		
FT	Modified-site	/label= HYP		
FT	Modified-site	9		
FT	Modified-site	/label= HYP		
FT	Modified-site	12		
FT	Modified-site	/label= HYP		
FT	Modified-site	15		
FT	Modified-site	/label= HYP		
FT	Modified-site	18		
FT	Modified-site	/label= HYP		
FT	Modified-site	21		
FT	Modified-site	/label= HYP		
FT	Modified-site	24		
FT	Peptide	/label= Hyp		
FT	Peptide	25..39		
FT	Modified-site	40		
FT	Modified-site	/note= "linked via peptide bond to Pro39 of peptide sequence"		
FT	Modified-site	41		
FT	Modified-site	/note= "linked via peptide bond to Pro39 of peptide comprising residues 1-39 of present sequence"		
FT	US5576419-A.			
PN	PD	26-JUL-2001.		
XX	PD	26-DEC-2000; 2000WO-US34263.		
PD	XX			
XX	PR	21-JAN-2000; 2000US-0488725.		
PP	PR	25-APR-2000; 2000US-052317.		
PN	PR	09-JUL-2000; 2000US-0588042.		
XX	PR	19-JUL-2000; 2000US-060312.		
PD	PR	01-AUG-2000; 2000US-053450.		
XX	PR	14-SEP-2000; 2000US-062191.		
PP	PR	19-OCT-2000; 2000US-0593036.		
PN	PR	29-NOV-2000; 2000US-0727344.		
XX	PA	(HYSE-) HYSEQ INC.		
XX	PI	Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;		
PT	PI	Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;		
XX	PI	Zhao QA, Zhou P, Goodrich R, Drmanac RT;		
DR	DR	WPI; 2001-442253/47.		
XX	DR	NP-SDB; AAI60019.		
XX	CC	The present sequence is that of triple helical branched polypeptide - under O-glycosidic bond and deuterium label retaining conditions, which may contain collagen cell adhesion sequences, useful for studying structure and biological activity of collagen	CC	the encoded polypeptides (AAW38642-AAW42213) with nootropic, immunosuppressant and cytostatic activity. The polynucleotides are useful in gene therapy. A composition containing a polypeptide or polynucleotide of the invention may be used to treat diseases of the peripheral nervous system, such as peripheral nervous injuries, peripheral neuropathy and localised neuropathies and central nervous system diseases, such as
PT	CC	Claim 48; column 27; 25pp; English.	CC	
XX	PS	WPI; 1997-011309/01.	CC	
XX	PR	Solid phase synthesis of triple-helical branched polypeptide - under O-glycosidic bond and deuterium label retaining conditions, which may contain collagen cell adhesion sequences, useful for studying structure and biological activity of collagen	CC	
PT	CC	The present sequence is that of triple helical polypeptide 2 (THP-2), which includes a collagen cell adhesion sequence, known to promote adhesion and migration of various cells (residues 1263-1277 of the collagen alpha1(IV) chain), in its triple helix inducing region. THP-2 was prepared by solid phase synthesis under O-glycosidic bond and deuterium label retaining conditions, without	CC	
XX	PS	Claim 48; column 27; 25pp; English.	CC	
XX	PS	The present sequence is that of triple helical polypeptide 2 (THP-2), which includes a collagen cell adhesion sequence, known to promote adhesion and migration of various cells (residues 1263-1277 of the collagen alpha1(IV) chain), in its triple helix inducing region. THP-2 was prepared by solid phase synthesis under O-glycosidic bond and deuterium label retaining conditions, without	CC	
XX	PS	Novel nucleic acids and polypeptides, useful for treating disorders such as central nervous system injuries.	CC	
XX	PS	Example 2; SEQ ID NO 5794; 10078pp; English.	CC	
XX	PS	The invention relates to human nucleic acids (AA157798-AA161369) and	CC	
XX	DR	the encoded polypeptides (AAW38642-AAW42213) with nootropic, immunosuppressant and cytostatic activity. The polynucleotides are useful in gene therapy. A composition containing a polypeptide or polynucleotide of the invention may be used to treat diseases of the peripheral nervous system, such as peripheral nervous injuries, peripheral neuropathy and localised neuropathies and central nervous system diseases, such as	CC	

CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager Syndrome. Other uses include the utilisation of the activities such as: immune system suppression, CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic and thrombolytic activity, cancer diagnosis and therapy, drug screening, assays for receptor activity, arthritis and inflammation, leukaemias and C.N.S disorders.

Note: The sequence data for this patent did not form part of the printed specification.

SQ sequence 1669 AA;

Query Match 100.0%; Score 92; DB 22; Length 1669;
Best Local Similarity 100.0%; Pred. No. 7.5e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVKGDKGNPGWPGAP 15
Db 1263 GVKGDKGNPGWPGAP 1277

RESULT 15

AAM3907 standard; Protein; 1672 AA.
XX AAM3907;
XX DT 22-OCT-2001 (first entry)

DE Human polypeptide SEQ ID NO 2222.
XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer; peripheral nervous system; neuropathy; central nervous system; CNS; KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic; amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic; chemokinetic; thrombolytic; drug screening; arthritis; inflammation; leukæmia; KW Homo sapiens.
XX OS WO200153312-A1.
XX PD 26-JUL-2001.
XX PR 26-DEC-2000; 2000WO-US34263.

XX PR 21-JAN-2000; 2000US-0488725.
PR 25-JPR-2000; 2000US-055317.
PR 09-JUL-2000; 2000US-059042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.

XX (HYSE-) BYSEQ INC.
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Oian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Dimancic RT;
DR WPI; 2001-442253/47.
N-PSDB; AA158233.

XX Novel nucleic acids and polypeptides, useful for treating disorders such as central nervous system injuries -
PT Example 4; SEQ ID NO 2222; 10078pp; English.
XX The invention relates to human nucleic acids (AA157798-AA161369) and the encoded polypeptides (AAm3862-AAm42213) with nootropic, immunosuppressant and cytostatic activity. The polynucleotides are useful in gene therapy. A composition containing a polypeptide or polynucleotide sequence 1669 AA;

XX (HYSE-) BYSEQ INC.
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Oian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Dimancic RT;
DR WPI; 2001-442253/47.
N-PSDB; AA158233.

XX Novel nucleic acids and polypeptides, useful for treating disorders such as central nervous system injuries -
PT Example 4; SEQ ID NO 2222; 10078pp; English.
XX The invention relates to human nucleic acids (AA157798-AA161369) and the encoded polypeptides (AAm3862-AAm42213) with nootropic, immunosuppressant and cytostatic activity. The polynucleotides are useful in gene therapy. A composition containing a polypeptide or polynucleotide sequence 1669 AA;

CC of the invention may be used to treat diseases of the peripheral nervous system, such as peripheral nervous injuries, peripheral neuropathy and localised neuropathies and central nervous system diseases, such as Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager Syndrome. Other uses include the utilisation of the activities such as: immune system suppression, Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic and thrombolytic activity, cancer diagnosis and therapy, drug screening, assays for receptor activity, arthritis and inflammation, leukaemias and C.N.S disorders.

Note: The sequence data for this Patent did not form part of the printed specification.

SQ Sequence 1672 AA;

Query Match 100.0%; Score 92; DB 22; Length 1672;
Best Local Similarity 100.0%; Pred. No. 7.5e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVKGDKGNPGWPGAP 15
Db 1265 GVKGDKGNPGWPGAP 1280

RESULT 16

XX AAB50169;
ID AAB50169 standard; Protein; 406 AA.
XX AAB50169;
AC AAB50169;
XX DT 14-MAR-2001 (first entry)

XX DE Lung cancer associated polypeptide sequence SEQ ID 507.
XX Human; lung cancer associated protein; neuroprotective; cytostatic; cardioactive; immunomodulatory; muscular active; vulnerable; gastrointestinal; nephrotropic; antiinfective; gynecological; antibacterial; diagnosis; neural disorder; immune disorder; reproductive; proliferative disorder; wound healing; infectious disease.
KW Homo sapiens.
XX OS WO200055180-A2.
PN XX
PN 21-SEP-2000.
XX PR 08-MAR-2000; 2000WO-US05918.
XX PR 12-MAR-1999; 9905-0124270.
XX PA (HUWA-) HUMAN GENOME SCI INC.
PA (ROSE/) ROSEN C A.
XX PI Ruben SM;
XX DR WPI; 2000-587514/55.
DR N-PSDB; AA1F18045.

XX Lung cancer associated gene sequences, referred to as lung cancer associated antigens, useful for treatment, prevention, and diagnosis of disorders such as lung cancer -
PT Claim 11; Page 996-998; 1425pp; English.
XX Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer associated proteins represented in AAB58106 - AAB5848. Lung cancer associated proteins and polynucleotide sequences, their agonists, and antagonists may have neuroprotective; cytotoxic; cardioactive; immunomodulatory; muscular active general; vulnary; gastrointestinal; general; nephrotoxic; antiinfective; gynaecological; or antibacterial activity. The invention also includes antibodies specific for the protein or polynucleotide sequences. The lung cancer associated polynucleotide sequences may be used for detection of lung cancer,

CC chromosome identification, as chromosome markers, and for numerous other
 CC diagnostic or research purposes. The proteins may be used to treat
 CC disorders such as neural, immune, muscular, reproductive,
 CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
 CC disorders. The proteins may also be used in the treatment of wounds and
 CC infectious diseases. Polynucleotide sequences AAF1425 - AAF1843 and
 CC peptide AAB58549 are used in the course of the invention for the
 CC identification and characterisation of the polynucleotide and protein
 CC sequences.

SQ Sequence 406 AA;

Query Match 93.5%; Score 86; DB 21; Length 406;
 Best Local Similarity 100.0%; Pred. No. 0.00013; Mismatches 0; Indels 0; Gaps 0;
 Matches 14; Conservative 0; MisMatches 0; OS

OY 2 VKGDKGNGPGWPGAP 15
 DR 1 VKGDKGNGPGWPGAP 14

RESULT 17

AR23873 ID AR23873 standard; Protein: 772 AA.
 XX AC AAR23873;
 XX DT 25-NOV-1992 (first entry)
 XX DE Human alpha 5 (IV) of type IV collagen.
 XX KW Mutations: Alport's syndrome; basement membranes; diabetes mellitus.
 XX OS Homo sapiens.

FH Key Location/Qualifiers

FT Misc-difference 43..47
 FT /note= "interruption in Gly-X-Y sequence"
 FT Misc-difference 159..160

FT /note= "interruption in Gly-X-Y sequence"
 FT Misc-difference 275..277

FT Misc-difference 334..336
 FT /note= "interruption in Gly-X-Y sequence"
 FT Misc-difference 456..458

FT /note= "interruption in Gly-X-Y sequence"
 PN US5114140-A.

XX 19-MAY-1992.
 XX PR 07-JUL-1989; 89US-0377238.

PR 07-JUL-1989; 89US-0377238.
 XX PA (TRYG/) TRYGGVASON K.

PI Hostikka SL, Tryggvason K;
 XX WPI: 1992-192174/23.

DR N-PSDB; AAQ24551.
 XX Isolation of DNA encoding alpha-5(IV)polypeptide of type IV
 PT collagen - to detect mutations in genes for alpha-5(IV) chain
 PT which produce genetic or acquired basement membrane disorders
 e.g. Alport's syndrome

XX Disclosure; Fig 2; 14pp; English.
 XX The sequence is that of the alpha 5(IV) polypeptide chain of human
 CC type IV collagen, the major component of basement membranes. The
 protein contains the Gly-X-Y repeat coding sequence typical for
 CC collagenous proteins at one end and a typical NC-domain coding

CC sequence at the other end. The sequence can be used to detect
 CC mutations in individual genes specific for this chain which can,
 CC directly or indirectly, produce several human diseases. It can
 CC also be used to determine genetic, e.g. Alport's syndrome, or
 CC acquired e.g. diabetes mellitus, disorders of the basement membrane,
 CC and as probes or antibodies against these nucleotide sequences.
 CC Gene fragments generated through amplifications from human genomic
 CC or cloned DNA can also be used for detection and analysis of genes.

SQ Sequence 772 AA;

Query Match 75.0%; Score 69; DB 13; Length 772;
 Best Local Similarity 73.3%; Pred. No. 0.074; Mismatches 2; Indels 0; Gaps 0;
 Matches 11; Conservative 2; MisMatches 0; OS

OY 1 GVKGDKGNPGWPGAP 15
 DR 366 GVKGEKGNPGGCGLP 380

RESULT 18

AAW09643 ID AAW09643 standard; Protein: 772 AA.
 XX AC AAW09643;
 XX DT 16-JUN-1997 (first entry)
 XX DE Human type IV collagen alpha-5.
 XX KW Collagen alpha5(IV); basement membrane; Alport's syndrome;
 KW nephritis; kidney; renal failure; antibody; diagnosis; COL4A5 gene;
 X chromosome.
 XX OS Homo sapiens.

FH Key Location/Qualifiers

FT Domain 1..543
 FT /label= Collagenous domain contains Gly-X-Y

FT /note= "collagenous domain contains Gly-X-Y
 tripeptide repeat; interrupted at positions 43-47, 159-160, 275-276, 334-335,
 456-459".

FT Domain 544..772
 FT /label= Non-collagenous_domain 742..751

FT Peptide 742..751
 FT /label= Immunogenic_peptide
 FT /note= "peptide used to raise diagnostic antibodies (Claim 1)"

FT XX US5593900-A.
 XX PD 14-JAN-1997.

XX PR 07-JUL-1989; 89US-0377238.

XX PA (HOST/ HOSIINKA S L.
 PA (HOST/ HOHTYA M.

PA (TRYG/) TRYGGVASON K.

XX PI Hostikka SL, Hoyhtya M, Tryggvason K;

XX DR WPI: 1997-099481/09.

DR N-PSDB; AAU47812.

XX PT New antibodies specific for human type IV collagen alpha5 chain -
 PT used to detect absence of this chain in patients with renal failure

CC Disclosure; Fig 2A-2B; 12pp; English.

xx The amino acid sequence of a portion (AAW09643) of the previously
 cc unknown human type IV collagen chain, alpha5(IV), was deduced from
 cc cDNA clones (see also AAT47812), obid, using probes based on
 cc conserved sequences of human alpha1(IV) and alpha2(IV) collagen
 chains and of the Drosophila alpha1(IV) chain. It includes a
 complete non-collagenous domain that shows 83% identity with that
 cc of alpha1(IV) and 63% with that of the alpha2(IV) chain. Mutations
 in the alpha5(IV) gene (COL4A5) are associated with Alport's
 cc syndrome. Antibodies raised against a peptide (see also AAW09644)
 specific to alpha5(IV) can be used in the diagnosis of basement
 cc membrane disorders such as Alport's syndrome.

xx Sequence 772 AA;

SQ

Query Match	Score	DB	Length
Best Local Similarity	75.0%	18	772
Matches	Pred. No.	Mismatches	Indels
11;	0.074;	2;	0;
			Gaps
			0;

OY 1 GVKGDKGNPGNPGAP 15
 Db 366 GIKGEKGNPGQPGGLP 380

RESULT 19

ABG04839

ID ABG04839 standard; Protein: 1685 AA.

XX

AC ABG04839;

XX

DT 13-FEB-2002 (first entry)

DE Novel human diagnostic protein #4830.

XX

KW Human; chromosome mapping; gene mapping; gene therapy; forensic; food supplement; medical imaging; diagnostic; genetic disorder.

XX

Homo sapiens.

OS

XX

PN WO200175067-A2.

XX

PD 11-OCT-2001.

XX

PR 30-MAR-2001; 2001WO-US08631.

XX

PR 31-MAR-2000; 2000US-0540217.

XX

PR 23-AUG-2000; 2000US-0649167.

XX

PA (HYSE-) HYSEQ INC.

XX

PI Drmanac RT, Liu C, Tang YT;

XX

DR WPI: 2001-639362/73.

XX

N-PSDB; AAS69026.

XX

New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity

PT

CC

The invention relates to isolated polynucleotide (I) and

CC

polypeptide (II) sequences. (I) is useful as hybridisation probes,

CC

polymerase chain reaction (PCR) primers, oligomers, and for chromosome

CC

and gene mapping, and in recombinant production of (II). The

CC

polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques for identifying expressed genes. (I) is useful as hybridisation probes, for restoring normal activity of (II) or to treat disease states involving

CC

(II). (II) is useful for generating antibodies against it, detecting or

CC

quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical

CC

imaging of sites expressing (II). (I) and (II) are useful for treating

CC

disorders involving aberrant protein expression or biological activity.

CC

The polypeptide and polynucleotide sequences have applications in

CC

diagnostics, forensics, gene mapping, identification of mutations

CC

responsible for genetic disorders or other traits to assess biodiversity

CC

and to produce other types of data and products dependent on DNA and

CC

amino acid sequences.

CC

ABG0010-ABG30377 represents novel human

CC

diagnostic amino acid sequences of the invention.

CC

Note: The sequence data for this patent did not appear in the printed

CC

specification, but was obtained in electronic format directly from WIPO

CC

at [ftp://www.wipo.int/pdb/published_pct_sequences](http://www.wipo.int/pdb/published_pct_sequences).

xx Sequence 1693 AA;

SQ

Query Match	Score	DB	Length
Best Local Similarity	75.0%	22	1693
Matches	Pred. No.	Mismatches	Indels
11;	0.17;	2;	0;
			Gaps
			0;

OY 1 GVKGDKGNPGNPGAP 15
 Db 1279 GIKGEKGNPGQPGGLP 1293

RESULT 20

ABG15619

ID ABG15619 standard; Protein: 1693 AA.

XX

AC ABG15619;

XX

DT 18-FEB-2002 (first entry)

DE Novel human diagnostic protein #15610.

XX

KW Human; chromosome mapping; gene mapping; gene therapy; forensic; food supplement; medical imaging; diagnostic; genetic disorder.

XX

OS Homo sapiens.

XX

PN WO200175067-A2.

XX

PD 11-OCT-2001.

XX

PR 30-MAR-2001; 2001WO-US08631.

XX

PR 31-MAR-2000; 2000US-0540217.

XX

PR 23-AUG-2000; 2000US-0649167.

XX

PA (HYSE-) HYSEQ INC.

XX

PI Drmanac RT, Liu C, Tang YT;

XX

DR WPI: 2001-639362/73.

XX

N-PDB; AAS79806.

XX

New isolated polynucleotide and encoded polypeptides, useful in

PT

diagnostics, forensics, gene mapping, identification of mutations

PT

responsible for genetic disorders or other traits and to assess

PT

biodiversity

XX

Claim 20; SEQ ID No 35198; 103pp; English.

XX

The invention relates to isolated polynucleotide (I) and

CC

polypeptide (II) sequences. (I) is useful as hybridisation probes,

CC

polymerase chain reaction (PCR) primers, oligomers, and for chromosome

CC

and gene mapping, and in recombinant production of (II). The

CC

polynucleotides are also used in diagnostics as expressed sequence tags

CC

for identifying expressed genes. (I) is useful in gene therapy techniques

CC

for restoring normal activity of (II) or to treat disease states involving

CC

(II). (II) is useful for generating antibodies against it, detecting or

CC

quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical

CC

imaging of sites expressing (II). (I) and (II) are useful for treating

CC

disorders involving aberrant protein expression or biological activity.

CC

The polypeptide and polynucleotide sequences have applications in

CC

diagnostics, forensics, gene mapping, identification of mutations

CC

responsible for genetic disorders or other traits to assess biodiversity

CC

and to produce other types of data and products dependent on DNA and

CC

amino acid sequences.

CC

ABG0010-ABG30377 represents novel human

CC

diagnostic amino acid sequences of the invention.

CC

Note: The sequence data for this patent did not appear in the printed

CC

specification, but was obtained in electronic format directly from WIPO

CC

at [ftp://www.wipo.int/pdb/published_pct_sequences](http://www.wipo.int/pdb/published_pct_sequences).

CC

The polypeptide and polynucleotide sequences have applications in

CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG0010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at [ftp://wipo.int/pub/published_pct_sequences](http://wipo.int/pub/published_pct_sequences).

SQ Sequence 1693 AA; Query Match 75.0%; Score 69; DB 22; Length 1693;
 Best Local Similarity 73.3%; Pred. No. 0.17; Matches 11; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 GVKGDKGNPGNPGAP 15
 | : | : | : | : |
 Db 1287 GIKGEKGNPGPGLP 1301

RESULT 21

AAB58180
 ID AAB58180 standard; Protein: 430 AA.

AC AAB58180;
 XX
 DT 14-MAR-2001 (first entry)

DE Lung cancer associated polypeptide sequence SEQ ID 518.

XX Human: lung cancer associated protein; neuroprotective; cytostatic;
 KW cardiotropic; immunomodulatory; muscular active; vulnery;
 KW gastrointestinal; nephrotoxic; antinfective; gynecological;
 KW antibacterial; diagnosis; neural disorder; immune disorder; reproductive;
 KW proliferative disorder; wound healing; infectious disease.
 OS Homo sapiens.
 XX
 PN WO20055180-A2.

PD 21-SEP-2000.

XX
 PF 08-MAR-2000; 2000WO-US05918.

XX
 PR 12-MAR-1999; 99US-0124270.

XX
 PA (HUMA-) HUMAN GENOME SCI INC.

PA (ROSE/-) ROSEN C A.

XX
 PT Ruben SM;

XX
 DR WPI; 2000-587514/55.

DR N-PSDB; AAF18056.

XX
 PT Lung cancer associated gene sequences, referred to as lung cancer
 PT antigens, useful for treatment, prevention, and diagnosis of disorders
 PT such as lung cancer -

XX
 PS Claim 11; Page 1008-1010; 1425pp; English.

XX
 CC Polynucleotide sequences AAF17982 - AAF8424 encode human lung cancer
 CC associated proteins represented in AAB58106 - AAB58548. Lung cancer
 CC associated proteins and polynucleotide sequences, their agonists, and
 CC antagonists may have neuroprotective; cytostatic; cardioactive;
 CC immunomodulatory; muscular active general; pulmonary; gastrointestinal
 CC general; nephrotoxic; antinfective; gynecological; or antibacterial
 CC activity. The invention also includes antibodies specific for the
 CC protein or polynucleotide sequences. The lung cancer associated
 CC polynucleotide sequences may be used for detection of lung cancer,
 CC chromosome identification, as chromosome markers, and for numerous other
 CC diagnostic or research purposes. The proteins may be used to treat
 CC disorders such as neural, immune, muscular, reproductive,
 CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative

CC disorders. The proteins may also be used in the treatment of wounds and
 CC infectious diseases. Polynucleotide sequences AAF8425 - AAF18433 and
 CC peptide AB5849 are used in the course of the invention for the
 CC identification and characterisation of the polynucleotide and protein
 CC sequences.

SQ Sequence 430 AA; Query Match 69.6%; Score 64; DB 21; Length 430;
 Best Local Similarity 73.3%; Pred. No. 0.21; Matches 11; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 GVKGDKGNPGNPGAP 15
 | : | : | : | : |
 Db 86 GPKGPKGDPGPGAP 100

RESULT 22

ABB68977
 ID ABB68977 standard; Protein: 224 AA.

AC ABB68977;
 XX
 DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 33723.

KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.

OS Drosophila melanogaster.

XX
 PN WO200171042-A2.

XX
 PR 11-JUL-2000; 2000US-0614150.

XX
 PD 27-SEP-2001.

XX
 PF 23-MAR-2001; 2001WO-US09231.

XX
 PR 23-MAR-2000; 2000US-191637P.

XX
 PR 11-JUL-2000; 2000US-0614150.

XX
 PA (PEKE) PE CORP NY.

XX
 PI Venter JC, Adams M, Li PWD, Myers EW;

XX
 DR WPI; 2001-656860/75.

DR N-PSDB; ABL13080.

XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -

XX
 PS Disclosure; SEQ ID NO 33723; 21pp + Sequence Listing; English.

XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL3011), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (ABB57737-ABB2072).

CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at [ftp://wipo.int/pub/published_pct_sequences](http://wipo.int/pub/published_pct_sequences).

XX
 SQ Sequence 224 AA;
 CC Query Match 66.3%; Score 61; DB 22; Length 224;
 CC Best Local Similarity 66.7%; Pred. No. 0.29;
 CC Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 1 GVKGDKGNPGNPGAP 15

Db 118 | :||||:||| ||| |
 118 GPRGDKGHPOMPGrP 132

RESULT 23
 AAU16066 ID AAU16066 standard; protein; 429 AA.
 XX AC AAU16066;
 XX DT 07-NOV-2001 (first entry)
 XX DE Human novel secreted protein, Seq ID 1019.
 XX KW Human; immunosuppressive; antiarthritic; antirheumatic;
 KW cytostatic; cardiotropic; vasotropic; cerebroprotective; nootropic;
 KW neuroprotective; antibacterial; viricide; fungicide; ophthalmological;
 KW vulnerability; secreted protein; rheumatoid arthritis;
 KW hyperproliferative disorder; cardiovascular disorder; cardiac arrest;
 KW cerebrovascular disorder; cerebral ischaemia; angiogenesis;
 KW nervous system disorder; Alzheimer's disease; infection; ocular disorder;
 KW corneal infection; wound healing; epithelial cell proliferation;
 KW skin ageing; food additive; preservative; antiproliferative.
 XX OS Homo sapiens.
 XX PN WO200155322-A2.
 XX PD 02-AUG-2001.
 XX
 PF 17-JAN-2001; 2001WO-US01341.
 XX PR 31-JAN-2000; 2000US-0179065.
 PR 04-FEB-2000; 2000US-0180628.
 PR 24-FEB-2000; 2000US-0184664.
 PR 02-MAR-2000; 2000US-0186350.
 PR 16-MAR-2000; 2000US-0189874.
 PR 17-MAR-2000; 2000US-0190076.
 PR 18-APR-2000; 2000US-198123.
 PR 19-MAY-2000; 2000US-0205515.
 PR 07-JUN-2000; 2000US-0209467.
 PR 28-JUN-2000; 2000US-0214886.
 PR 07-JUL-2000; 2000US-0215135.
 PR 07-JUL-2000; 2000US-0216647.
 PR 11-JUL-2000; 2000US-0216880.
 PR 11-JUL-2000; 2000US-0217487.
 PR 14-JUL-2000; 2000US-0218290.
 PR 26-JUL-2000; 2000US-0220963.
 PR 14-AUG-2000; 2000US-0224518.
 PR 14-AUG-2000; 2000US-0224519.
 PR 14-AUG-2000; 2000US-0225213.
 PR 14-AUG-2000; 2000US-0225214.
 PR 14-AUG-2000; 2000US-0225266.
 PR 14-AUG-2000; 2000US-0225267.
 PR 14-AUG-2000; 2000US-0225268.
 PR 22-AUG-2000; 2000US-0225447.
 PR 14-AUG-2000; 2000US-0225757.
 PR 14-AUG-2000; 2000US-022718.
 PR 23-AUG-2000; 2000US-0227009.
 PR 18-AUG-2000; 2000US-0228924.
 PR 01-SEP-2000; 2000US-0229287.
 PR 01-SEP-2000; 2000US-0229343.
 PR 01-SEP-2000; 2000US-0229344.
 PR 05-SEP-2000; 2000US-0229509.

PR 05-SEP-2000; 2000US-0229513.
 PR 06-SEP-2000; 2000US-0230437.
 PR 06-SEP-2000; 2000US-0230438.
 PR 08-SEP-2000; 2000US-0231242.
 PR 08-SEP-2000; 2000US-0231243.
 PR 08-SEP-2000; 2000US-0231413.
 PR 08-SEP-2000; 2000US-0231414.
 PR 08-SEP-2000; 2000US-0232080.
 PR 08-SEP-2000; 2000US-0232081.
 PR 12-SEP-2000; 2000US-0231968.
 PR 12-SEP-2000; 2000US-0232397.
 PR 14-SEP-2000; 2000US-0232398.
 PR 14-SEP-2000; 2000US-0232399.
 PR 14-SEP-2000; 2000US-0232423.
 PR 14-SEP-2000; 2000US-0232401.
 PR 14-SEP-2000; 2000US-0233063.
 PR 14-SEP-2000; 2000US-0233064.
 PR 14-SEP-2000; 2000US-0233065.
 PR 21-SEP-2000; 2000US-0234274.
 PR 21-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234997.
 PR 25-SEP-2000; 2000US-0234998.
 PR 26-SEP-2000; 2000US-0235484.
 PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235836.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 02-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-023935.
 PR 13-OCT-2000; 2000US-023937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241121.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 01-NOV-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0244674.
 PR 08-NOV-2000; 2000US-0244675.
 PR 08-NOV-2000; 2000US-0244676.
 PR 08-NOV-2000; 2000US-0244677.
 PR 08-NOV-2000; 2000US-0244678.
 PR 08-NOV-2000; 2000US-0244679.
 PR 08-NOV-2000; 2000US-0244679.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.

PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249287.
 PR 17-NOV-2000; 2000US-0249299.
 PR 01-DEC-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250301.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0255678.
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX Rosen CA, Barash SC, Ruben SM;
 PT WPI; 2001-488783/53.
 DR N-PSDB; AAS26053.
 XX
 PT New nucleic acid molecules encoding 461 human secreted proteins for
 PT diagnosing, preventing, treating or ameliorating medical conditions and
 PT used as food additives or preservatives -
 XX
 PS Claim 11; SEQ ID NO 1019; 980PP; English.
 XX
 CC The invention relates to isolated nucleic acid molecules and their
 CC encoded secreted proteins. The nucleic acids and proteins are used to
 CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
 CC rabbits, goats, horses, cats, dogs, chickens or sheep. They
 CC are also used in diagnosing a pathological condition or susceptibility
 CC to a pathological condition. Antibodies to the proteins can also
 CC be used in alleviating symptoms associated with the disorders and in
 CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked
 CC immunosorbant assays (ELISA). Disorders which are diagnosed or treated
 CC include autoimmune diseases e.g. rheumatoid arthritis,
 CC hyperproliferative disorders e.g. neoplasms of the breast or liver,
 CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
 CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.
 CC Alzheimer's disease, infections caused by bacteria, viruses and fungi.
 CC and ocular disorders e.g. corneal infection, and many other
 CC disorders listed in the specification. The polypeptides can also
 CC be used to aid wound healing and epithelial cell proliferation, to
 CC prevent skin aging due to sunburn, to maintain organs before
 CC transplantation, for supporting cell culture of primary tissues, to
 CC regenerate tissues and in chemotaxis. The polypeptides can also be used
 CC as a food additive or preservative to increase or decrease storage
 CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,
 CC minerals, cofactors and other nutritional components. The present
 CC sequence represents a novel secreted protein of the invention.

Query Match Score 61; DB 22; Length 429;
 Best Local Similarity 66.3%; Pred. No. 0.58;
 Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 1 GVKGDKGNGPWNQGP 15
 Db 23 GAKGEKGKGNAGFPGLP 37

RESULT 24
 ID AAW40114 standard; Protein; 532 AA.

XX AC AAW40114;
 XX DT 03-JUN-1998 (first entry)
 XX DE Human alpha-1(IV) collagen protein.
 XX KW COL4A6 gene; type IV collagen protein; alpha-6(IV); alpha-1(IV);
 KW Alports syndrome; leiomyomatosis; diagnosis; gene therapy; antibody.
 XX OS Homo sapiens.

XX Location/Qualifiers
 FT Key 163..171
 FT Region /label= I
 FT /note= "This region interrupts a collagenous repeat"
 FT 224..225
 FT Region /label= II
 FT /note= "This region interrupts a collagenous repeat"
 FT 244..257
 FT Region /label= III
 FT /note= "This region interrupts a collagenous repeat"
 FT 273..274
 FT Region /label= IV
 FT /note= "This region interrupts a collagenous repeat"
 FT 335..337
 FT Region /label= V
 FT /note= "This region interrupts a collagenous repeat"
 FT 377..378
 FT Region /label= VI
 FT /note= "This region interrupts a collagenous repeat"
 FT 403..404
 FT Region /label= VII
 FT /note= "This region interrupts a collagenous repeat"
 FT 429..236
 FT Region /label= VIII
 FT /note= "This region interrupts a collagenous repeat"
 FT 467..473
 FT Region /label= IX
 FT /note= "This region interrupts a collagenous repeat"
 XX PN US5731192-A.
 XX PD 24-MAR-1998.
 XX PF 23-JUN-1995; 95US-0494168.
 XX PR 27-AUG-1993; 93US-0112465.
 XX PA (UYYA) UNIV YALE.
 XX PI Reeder ST, Zhou J;
 XX DR WPI; 1998-216495/19.
 XX
 PT Nucleic acid encoding human alpha-6(IV) collagen - useful for, e.g.
 PT diagnosis or gene therapy of Alport's disease
 XX
 PS Example 1; FIG 3; 43pp; English.
 XX
 CC This sequence is the human alpha-1(IV) collagen protein which is used
 CC to analyse a novel human alpha-6(IV) collagen protein encoded by the
 CC COL4A6 gene. The alpha-6(IV) protein can be used for diagnosis or gene
 CC therapy of diseases associated with collagen type IV pathology,
 CC especially Alport's syndrome and associated diffuse leiomyomatosis. The
 CC polypeptide may also be used for generating monoclonal or polyclonal
 CC antibodies having specificity for the alpha-6(IV) polypeptide especially
 CC an antibody that is not crossreactive with other collagen proteins
 CC including alpha-1(IV), alpha-2(IV) and alpha-5(IV) collagens.
 XX Sequence 532 AA;

Query Match Score 61; DB 19; Length 532;

Best Local Similarity 66.7%; Pred. No. 0.73; Matches 10; Conservative 3; Mismatches 2; Indels 0; Gaps 0; OSY 1 GVKGDKGNGWPGAP 15
 Query Match 66.3%; Score 61; DB 22; Length 1690; Best local Similarity 60.0%; Pred. No. 2.5; Matches 9; Conservative 4; Mismatches 2; Indels 0; Gaps 0; OSY 1 GVKGDKGNGWPGAP 15
 Db 184 GIOQDRGDPGLPULP 198

RESULT 25
 AAM23916 standard; Protein; 1690 AA.
 XX AC AAR1704;
 XX DT 17-OCT-1995 (first entry)
 XX DE Collagen alpha 1 (III) chain precursor.
 XX KW Collagen; antibody; immunoassay; metabolism; diagnosis; monitoring; disorder; osteoporosis; metastatic progression; Paget's disease; hyperthyroid; bone; resorption; rheumatoid arthritis; osteoarthritis; vasculitis syndrome.
 XX OS Homo sapiens.
 XX PN WO9508115-A.
 XX PD 23-MAR-1995.
 XX PR 17-SEP-1994; 94WO-DK000348.
 XX PA (OSTE-) OSTEOmeter AS.
 XX PI Bonde M, Qvist P;
 XX DR WPI; 1995-131456/17.
 XX PT Assaying collagen fragments in body fluid by immunoassay - using antibodies raised against synthetic peptide(s) contg. potential crosslinking sites, to diagnose and monitor disorders of collagen metabolism, e.g. osteoporosis.
 XX CC Disclosure (Appendix A); Page 55; 87pp; English.
 XX PS Determination of collagen fragments in body fluids can be achieved by immunoassay using antibodies directed against synthetic peptides derived from collagen which contain sites of potential crosslinking. The method is used to diagnose and monitor treatment of disorders of collagen metabolism (degradation of type I collagen may indicate osteoporosis, metastatic progression, Paget's disease, hyperthyroidism or other conditions involving excessive bone resorption; degradation of type II collagen may indicate rheumatoid arthritis or osteoarthritis; and of type III collagen, vacuilitis syndrome). The method can also be used to assess the toxicity of a compound and to test drugs for their effect on collagen metabolism.
 XX SQ Sequence 1078 AA;

The present invention provides the protein and coding sequences of novel proteins from a variety of organisms, including human, dog, cat, horse, cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea urchin and tomato. These were derived from expressed sequence tags (ESTs) from the organism of interest. They can be used in diagnostics, forensics, gene mapping, identification of mutations, to assess biodiversity and for nutritional purposes. The present sequence is a protein of the invention.

XX Sequence 1690 AA;

Query Match 65.2%; Score 60; DB 16; Length 1078; Best Local Similarity 71.4%; Pred. No. 2.2; Matches 10; Conservative 3; Mismatches 1; Indels 0; Gaps 0; OSY 1 GVKGDKGNGWPGA 14
 Db 711 GVKGERGSPGGPA 724

RESULT 27
 AAV96125 standard; Peptide; 1078 AA.
 XX ID AAV96125
 XX AC AAV96125;
 XX DT 19-DEC-2000 (first entry)
 XX DE Collagen type III alpha-1.
 XX KW Collagen type III; vasculitis syndrome; assay; diagnosis.
 XX OS Homo sapiens.
 XX PN US6110689-A.

RESULT 26
 AAR71704 standard; protein; 1078 AA.
 XX

PD 29-AUG-2000.
 XX PT - contains primer to amplify portions of Type III procollagen DNA
 PF 04-NOV-1997; PT and detects mutation in standard procollagen DNA
 XX PR Disclosure: Fig 1A-F; 44pp; English.
 XX (OSTE-) OSTOMETER AS.
 XX CC Example 1 describes the determination of the presence of a mutation
 PA CC in the pro-alpha1(III) gene. Primers used in PCR are given in
 XX AAQ30834-48.
 PI Bonde M, Qvist P;
 XX PT
 DR WPI; 2000-586349/55.
 XX PT Assaying type I collagen fragments for diagnosing osteoporosis in
 PT postmenopausal woman, involves contacting body fluid with synthetic
 PT collagen peptide and antibody and quantifying by competitive binding
 XX assay -
 PS disclosure; Column 46-51; 41pp; English.
 XX
 CC The present sequence is that of human type III collagen alpha-1.
 CC The invention is based on the discovery of the presence of
 CC particular collagen fragments in body fluids of patients compared
 CC with those of healthy subjects. These fragments are generated
 CC upon collagen degradation and are partly characterised by the
 CC presence of potential sites for crosslinking. A method for
 CC assaying collagen fragments in a body fluid sample is based on the
 CC competitive binding to immunological binding partners of collagen
 CC fragments in the sample and of synthetic peptides derived from
 CC collagen and containing crosslinkable sites (see AAY96118-21). When
 CC considering the degradation of type III collagen, the assay can be
 CC used as a means of identifying the presence of vasculitis syndrome.
 XX
 SQ Sequence 1078 AA:

Query Match	65.2%	Score	60;	DB	21;	Length	1196;
Best Local Similarity	71.4%	Pred.	No.	2.4;			
Matches	10;	Conservative	3;	Mismatches	3;	Gaps	0;
Indels	1;						
Gaps	0;						

 Oy 1 GVKGDKGKPGWPPGA 14
 Db 711 GVKGERGSPGGPPGA 724
 XX
 RESULT 28
 AAR28316
 ID AAR28316 standard; Protein: 1196 AA.
 XX
 AC AAR28316;
 XX DT 24-MAR-1993 (first entry)
 XX DE Type III procollagen (prior art).
 XX KW Mutation; pro-alpha1(III); primer; PCR.
 XX OS Homo sapiens.
 XX PN WO20175177-A2.
 XX PN 11-OCT-2001.
 XX PD 03-APR-2001; 2001WO-US10947.
 XX PR 03-APR-2000; 2000US-194336P.
 XX PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 XX PI Morin PJ, Sherman-Baust CA, Pizer ES, Hough CD;
 XX DR WPI; 2001-626450/72.
 XX N-PSDB; ARA83117.
 XX
 PT Detecting and identifying ovarian tumor, identifying increased risk for
 PT developing ovarian cancer, and determining effectiveness of ovarian
 PT cancer treatment, by measuring expression level of ovarian tumor marker
 PT gene -
 XX PS Claim 23; Page 114-117; 140pp; English.
 XX
 PA (UYJE-) UNTV JEFFERSON THOMAS.
 XX
 PT The invention relates to methods for diagnosing and prognosing ovarian
 PT tumours in an individual via the detection and measurement of the
 PT expression of ovarian tumour marker genes (ABA83031-ABA83122, ABA83180,
 PT CC ABA83182 and ABA83184) or segments thereof (ABA83123-ABA83169, ABA83179,
 PT CC ABA8181 and ABA8183). The methods of the invention are useful for
 PT CC detecting an ovarian tumour in a patient, for identifying an individual
 CC at increased risk for developing ovarian cancer, in prognostic tests for
 CC assessing the relative severity of ovarian cancer, in tests for

CC monitoring a patient in remission from ovarian cancer and in tests for
 CC monitoring disease status in a patient being treated for ovarian cancer.
 CC The methods can additionally be used to identify a particular tumour as
 CC being an ovarian tumour (i.e., an epithelial ovarian tumour selected from
 CC serous cystadenoma, borderline serous tumour, serous cystadenocarcinoma,
 CC mucinous cystadenoma, borderline mucinous tumour, mucinous
 CC cystadenocarcinoma, endometrioid carcinoma, undifferentiated carcinoma,
 CC clear cell adenocarcinoma, cystadenofibroma, adenofibroma and Brenner
 CC tumour. The ovarian tumour marker genes of the invention were identified
 CC using SAGE (serial analysis of gene expression) and were found to be
 CC overexpressed in a broad variety of ovarian epithelial tumour cells
 relative to normal ovarian epithelial cells. The marker genes are
 CC implicated in immune response pathways, in the regulation of cell
 CC proliferation and in protein folding, and many of these are membrane-
 CC localised or secreted. In addition to their use as diagnostic and
 CC prognostic markers, the ovarian tumour marker genes or their encoded
 CC proteins may be used as therapeutic targets for the treatment and
 CC prevention of ovarian cancer. Sequences ABP50257-ABP50299 represent
 CC proteins encoded by ovarian tumour marker genes of the invention.

XX Sequence 1466 AA:
 SQ

Query Match 65.2%; Score 60; DB 22; Length 1466;
 Best Local Similarity 78.6%; Pred. No. 3;
 Matches 11; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 GVKGDKGNPGWPGA 14
 Db 738 GPKGDKGPGGCDA 751

RESULT 30

AAE02533 standard; Protein: 1466 AA.

XX AAE02533;

XX DT 10-AUG-2001 (first entry)

DE Bovine alpha(III) collagen #1.

XX Bovine; alpha(III) collagen; gelatin; cytostatic; viral infection;
 KW pharmaceutical; food industry; cosmetic; autoimmune disorder; vaccine;
 KW medical; arterial sealant; bone graft; dermal implant; haemostat; cancer;
 KW rheumatoid arthritis; beverage; photographic application.

OS Bos sp.

XX WO200134647-A2.

XX PD 17-MAY-2001.

XX PP 10-NOV-2000; 200000-US30792.

XX PR 12-NOV-1999; 95US-0439058.

XX PR 10-NOV-2000; 20000US-0439058.

PA (FIBR-) FIBROGEN INC.

PI Bell MP, Neff TB, Polarek JW, Seeley TW;

XX DR WPI; 2001-335911/35.

DR N-PSDB; AAD06574.

PT Novel isolated and purified bovine or porcine collagens and gelatins
 PT useful in medical, pharmaceutical, food and cosmetic industries, as
 PT vaccine, and for treating autoimmune disorders, infections and cancer

XX Example 2; Fig 4; 168pp; English.

CC The present sequence is bovine alpha(III) collagen. The present
 CC invention relates to recombinant synthesis of collagens and gelatins

CC derived from animals. Collagen is useful in medical pharmaceutical,
 CC food and cosmetic industries. Collagen is an important component of
 CC arterial sealants, bone grafts, drug delivery system, dermal implants,
 CC haemostats, and incontinence implants, and for treating autoimmune
 CC disorders such as rheumatoid arthritis. Collagen is useful in food
 CC products such as sausages casings, and in cosmetics or facial and skin
 CC formulations for treating viral infections, autoimmune diseases and
 CC cancer. Gelatin is useful in the manufacture or as a component of
 CC various pharmaceutical and medical devices and products, in food and
 CC beverage industries, in hair care and skin care products, as a glue or
 CC adhesive in various manufacturing processes, as a light-sensitive coating
 CC in various electronic devices, as photoresist base in photolithographic
 CC processes, in printing and photographic applications, in laboratory
 CC application, and as a component in various gels used for biochemical and
 CC electrophoretic analysis, including enzymographic gels.

XX Sequence 1466 AA:

Query Match 65.2%; Score 60; DB 22; Length 1466;
 Best Local Similarity 71.4%; Pred. No. 3;
 Matches 10; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 GVKGDKGNPGWPGA 14
 Db 858 GVKGGERGSPGGCDA 871

Search completed: November 1, 2002, 12:52:33
 Job time : 29.5 secs

OM protein - protein search, using SW mode1
 Run on: November 1, 2002, 12:52:42 ; Search time 10.5 Seconds
 (without alignments) 34.894 Million cell updates/sec

Title: US-09-529-691A-1
 Perfect score: 92
 Sequence: 1 GVKGDKGNPGNGPAGAP 15
 Scoring table: BLOSUM62
 Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
 Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
 Maximum Match 100%
 Listing first 45 summaries

Database : Issued_Patents_AA:*

1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep:*

2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep:*

3: /cgn2_6/ptodata/1/1aa/5A_COMB.pep:*

4: /cgn2_6/ptodata/1/1aa/5B_COMB.pep:*

5: /cgn2_6/ptodata/1/1aa/5C_COMB.pep:*

6: /cgn2_6/ptodata/1/1aa/5ACKFILES1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	92	100.0	15	3 US-08-394-748A-1	Sequence 1, Appli
2	92	100.0	15	3 US-08-394-748A-1	Sequence 1, Appli
3	92	100.0	15	5 PCT-US95-2478-1	Sequence 1, Appli
4	92	100.0	24	1 US-08-534-342-3	Sequence 3, Appli
5	92	100.0	24	1 US-08-675-140-3	Sequence 3, Appli
6	92	100.0	39	1 US-08-534-342-2	Sequence 2, Appli
7	92	100.0	39	1 US-08-675-140-2	Sequence 2, Appli
8	66	71.7	39	1 US-08-534-342-5	Sequence 5, Appli
9	66	71.7	39	1 US-08-534-342-6	Sequence 6, Appli
10	66	71.7	39	1 US-08-675-140-5	Sequence 5, Appli
11	66	71.7	39	1 US-08-675-140-6	Sequence 6, Appli
12	61	65.3	532	1 US-08-494-168-9	Sequence 9, Appli
13	60	65.2	1057	3 US-08-931-820-4	Sequence 4, Appli
14	60	65.2	1078	3 US-08-963-825-21	Sequence 21, Appli
15	60	65.2	1078	4 US-09-570-573-21	Sequence 21, Appli
16	60	65.2	1078	4 US-09-548-608-21	Sequence 21, Appli
17	59	64.1	471	2 US-08-399-889-24	Sequence 24, Appli
18	59	64.1	471	3 US-09-167-364-24	Sequence 24, Appli
19	59	64.1	471	4 US-09-439-889-72	Sequence 24, Appli
20	56	60.9	489	2 US-08-794-795-7	Sequence 7, Appli
21	56	60.9	489	4 US-09-249-200-7	Sequence 7, Appli
22	56	60.9	518	1 US-08-392-367B-2	Sequence 2, Appli
23	56	60.9	518	1 US-08-893-467A-2	Sequence 2, Appli
24	56	60.9	546	1 US-08-494-168-10	Sequence 10, Appli
25	55	59.8	171	3 US-09-011-735-2	Sequence 2, Appli
26	55	59.8	351	3 US-09-011-735-2	Sequence 1, Appli
27	55	59.8	1694	1 US-08-494-168-2	Sequence 2, Appli

ALIGMENTS

RESULT 1
 US-08-394-748A-1
 ; Sequence 1, Application US/08394748A
 ; Patent No. 6013628
 ; GENERAL INFORMATION:
 ; APPLICANT: Subritz, Amy P.N.
 ; APPLICANT: Furcht Leo T.
 ; APPLICANT: Balles, Mark
 ; APPLICANT: Greigerson, Dale S.
 ; APPLICANT: Agarwal, Anita
 ; APPLICANT: Wright, Marsha M.
 ; APPLICANT: Morali, Shobhana
 ; TITLE OF INVENTION: Method for Treating Conditions of the Eye
 ; TITLE OF INVENTION: Using Polypeptides
 ; NUMBER OF SEQUENCES: 16
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Merchant & Gould
 ; STREET: 3100 No. 6013628west Center
 ; CITY: Minneapolis
 ; STATE: MN
 ; COUNTRY: USA
 ; ZIP: 55402
 COMPUTER READABLE FORM:
 MEDIUM TYPE: FLOPPY disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/394,748A
 FILING DATE: 27-FEB-1995
 CLASSIFICATION: 514
 PRIORITY APPLICATION DATA:
 APPLICATION NUMBER: US 08/203,458
 FILING DATE: 28-FEB-1994
 ATTORNEY/AGENT INFORMATION:
 NAME: Carter, Charles G.
 REGISTRATION NUMBER: 35,093
 REFERENCE/DOCKET NUMBER: 600.307US01
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 612-332-5500
 TELEFAX: 612-332-9081
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 15 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 ORIGINAL SOURCE:
 STRAIN: IV-H1
 US-08-394-748A-1

Query Match 100.0%; Score 92; DB 3; Length 15;
 Best Local Similarity 100.0%; Pred. No. 2.1e-07; Mismatches 0;
 Matches 15; Conservative 0; Indels 0; Gaps 0;

Qy 1 GVKGDKGNGPWPAGAP 15
 Db 1 GVKGDKGNGPWPAGAP 15

RESULT 2

US-08-702-254A-1

Sequence 1, Application US/08702254A
 Patent No. 609683

GENERAL INFORMATION:

APPLICANT: FIELDS, GREGG B.
 TITLE OF INVENTION: SELF-ASSEMBLING AMPHIPHILES FOR CONSTRUCTION OF PEPTIDE SECONDARY STRUCTURES
 NUMBER OF SEQUENCES: 3
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: MUETTING, RAASCH & GEBHARDT, P.A.
 STREET: 119 NORTH FOURTH STREET
 CITY: MINNEAPOLIS
 STATE: MINNESOTA
 COUNTRY: USA
 ZIP: 55401

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US 08/703,458
 FILING DATE: 28-FEB-1994
 INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 15 amino acids
 TYPE: amino acid

TOPOLogy: linear

MOLECULE TYPE: peptide
 ORIGINAL SOURCE: STRAIN: IV-HI

RESULT 4

US-08-534-342-3

Sequence 3, Application US/08534342
 Patent No. 5576419

GENERAL INFORMATION:

APPLICANT: Fields, Gregg B.
 TITLE OF INVENTION: Mild Solid-Phase Synthesis of Aligned, Numbered Sequences
 NUMBER OF SEQUENCES: 13
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Merchant & Gould
 STREET: 3100 No. 5576419west Center
 CITY: Minneapolis
 STATE: MN
 COUNTRY: USA
 ZIP: 55402

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/534,342
 FILING DATE: 30-JUN-1993
 ATTORNEY/AGENT INFORMATION:

NAME: Mueting, Ann M.

REGISTRATION NUMBER: 33-977
 REFERENCE/DOCKET NUMBER: 600.253-US-01

TELECOMMUNICATION INFORMATION:

TELEPHONE: 612-333-5300
 TELEFAX: 612-332-9081

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 24 amino acids
 TYPE: amino acid

TOPOLogy: unknown

US-08-534-342-3

Query Match 100.0%; Score 92; DB 3; Length 15;
 Best Local Similarity 100.0%; Pred. No. 2.1e-07; Mismatches 0;
 Matches 15; Conservative 0; Indels 0; Gaps 0;

Qy 1 GVKGDKGNGPWPAGAP 15
 Db 1 GVKGDKGNGPWPAGAP 15

RESULT 3

PCT-US95-02478-1

Sequence 1, Application PC/TUS9502478
 GENERAL INFORMATION:

APPLICANT: Method for Treating Conditions
 TITLE OF INVENTION: Method for Treating Conditions
 NUMBER OF SEQUENCES: 16
 COMPUTER READABLE FORM:

COMPUTER: IBM PC compatible
 MEDIUM TYPE: Floppy disk
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25

RESULT 5
 US-08-675-140-3
 Sequence 3, Application US/08675140
 Patent No. 5726243
 GENERAL INFORMATION:
 APPLICANT: Fields, Gregg B.
 TITLE OF INVENTION: Mild Solid-Phase Synthesis of Aligned,
 NUMBER OF SEQUENCES: 13
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Merchant & Gould
 STREET: 3100 No. 5726243west Center
 CITY: Minneapolis
 STATE: MN
 COUNTRY: USA
 ZIP: 55402
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/675,140
 FILING DATE: 03-JUL-1996
 CLASSIFICATION: 525
 PRIORITY APPLICATION DATA:
 APPLICATION NUMBER: US 08/085,633
 FILING DATE: 30-JUN-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Mueeting, Ann M.
 REGISTRATION NUMBER: 33,977
 REFERENCE/DOCKET NUMBER: 600.253-US-01
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 612-332-5300
 TELEFAX: 612-332-9081
 INFORMATION FOR SEQ ID NO: 3:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 24 amino acids
 TYPE: amino acid
 TOPOLOGY: unknown
US-08-675-140-3

Query Match 100.0%; Score 92; DB 1; Length 24;
 Best Local Similarity 100.0%; Pred. No. 3,3e-07;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GVKGDKGNPGWPGAP 15
 Db 10 GVKGDKGNPGWPGAP 24

RESULT 6
 US-08-534-342-2
 Sequence 2, Application US/08534342
 Patent No. 556419
 GENERAL INFORMATION:
 APPLICANT: Fields, Gregg B.
 TITLE OF INVENTION: Mild Solid-Phase Synthesis of Aligned,
 NUMBER OF SEQUENCES: 13
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Merchant & Gould
 STREET: 3100 No. 5726243west Center
 CITY: Minneapolis
 STATE: MN
 COUNTRY: USA
 ZIP: 55402
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/675,140
 FILING DATE: 03-JUL-1996
 CLASSIFICATION: 525
 PRIORITY APPLICATION DATA:
 APPLICATION NUMBER: US 08/085,633
 FILING DATE: 30-JUN-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Mueeting, Ann M.
 REGISTRATION NUMBER: 33,977
 REFERENCE/DOCKET NUMBER: 600.253-US-01
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 612-332-5300
 TELEFAX: 612-332-9081
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 39 amino acids
 TYPE: amino acid
 TOPOLOGY: unknown
US-08-534-342-2

Query Match 100.0%; Score 92; DB 1; Length 39;
 Best Local Similarity 100.0%; Pred. No. 5,6e-07;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1 GVKGDKGNPGWPGAP 15	Db 25 GVKGDKGNPGWPGAP 39

RESULT 7
 US-08-675-140-2
 Sequence 2, Application US/08675140
 Patent No. 5726243
 GENERAL INFORMATION:
 APPLICANT: Fields, Gregg B.
 TITLE OF INVENTION: Mild Solid-Phase Synthesis of Aligned,
 NUMBER OF SEQUENCES: 13
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Merchant & Gould
 STREET: 3100 No. 5726243west Center
 CITY: Minneapolis
 STATE: MN
 COUNTRY: USA
 ZIP: 55402
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/675,140
 FILING DATE: 03-JUL-1996
 CLASSIFICATION: 525
 PRIORITY APPLICATION DATA:
 APPLICATION NUMBER: US 08/085,633
 FILING DATE: 30-JUN-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Mueeting, Ann M.
 REGISTRATION NUMBER: 33,977
 REFERENCE/DOCKET NUMBER: 600.253-US-01
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 612-332-5300
 TELEFAX: 612-332-9081
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 39 amino acids
 TYPE: amino acid
 TOPOLOGY: unknown
US-08-675-140-2

Query Match 100.0%; Score 92; DB 1; Length 39;
 Best Local Similarity 100.0%; Pred. No. 5,6e-07;

Matches 15; conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 8
US-08-534-342-5
; Sequence 5, Application US/08534342
; Patent No. 5576419
GENERAL INFORMATION:
APPLICANT: Fields, Gregg B.
TITLE OF INVENTION: Mild Solid-Phase Synthesis of Aligned,
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Merchant & Gould
STREET: 3100 No. 5576419west Center
CITY: Minneapolis
STATE: MN
COUNTRY: USA
ZIP: 55402

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/534,342
FILING DATE:
CLASSIFICATION: 530
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/085,633
FILING DATE: 30-JUN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Mueting, Ann M.
REGISTRATION NUMBER: 33,977
TELECOMMUNICATION INFORMATION:
TELEPHONE: 612-332-5000
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS: 5:
LENGTH: 39 amino acids
TYPE: amino acid
TOPOLOGY: unknown

US-08-534-342-6

Query Match 71.7%; Score 66; DB 1; Length 39;
Best Local Similarity 73.3%; Pred. No. 0.003; 0; Mismatches 4; Indels 0; Gaps 0;

RESULT 9
US-08-534-342-6
; Sequence 6, Application US/08534342
; Patent No. 5576419
GENERAL INFORMATION:
APPLICANT: Fields, Gregg B.
TITLE OF INVENTION: Mild Solid-Phase Synthesis of Aligned,
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Merchant & Gould
STREET: 3100 No. 5576419west Center
CITY: Minneapolis
STATE: MN
COUNTRY: USA
ZIP: 55402

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/675,140
FILING DATE: 03-JUL-1996
CLASSIFICATION: 525
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/085,633
FILING DATE: 30-JUN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Mueting, Ann M.
REGISTRATION NUMBER: 33,977
TELECOMMUNICATION INFORMATION:
TELEPHONE: 612-332-5300
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS: 5:
LENGTH: 39 amino acids
TYPE: amino acid
TOPOLOGY: unknown

US-08-675-140-5

Query Match 71.7%; Score 66; DB 1; Length 39;
 Best Local Similarity 73.3%; Pred. No. 0; Mismatches 0; Indels 4; Gaps 0;

Matches 11; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 GVKGDKGNPGPGAP 15
 Db 25 GVGXGXPGPAGP 39

RESULT 11

US-08-675-140-6

Sequence 6, Application US/08675140
 Patent No. 5726243
 GENERAL INFORMATION:
 APPLICANT: Fields, Gregg B.
 TITLE OF INVENTION: Mild Solid-Phase Synthesis of Aligned,
 NUMBER OF SEQUENCES: 13
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Merchant & Gould
 STREET: 3100 No. 5726243west Center
 CITY: Minneapolis
 STATE: MN
 COUNTRY: USA
 ZIP: 55402

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/675,140
 FILING DATE: 03-JUL-1996
 CLASSIFICATION: 525

PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/085,633
 FILING DATE: 30-JUN-1993

ATTORNEY/AGENT INFORMATION:
 NAME: Muetting, Ann M.
 REGISTRATION NUMBER: 33,977

TELECOMMUNICATION INFORMATION:
 TELEPHONE: 612-332-5300
 TELEX: 612-332-9081

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:
 LENGTH: 39
 TYPE: amino acid
 LENGTH: 39 amino acids
 TYPE: amino acid
 TOPOLOGY: unknown

US-08-675-140-6

Query Match 71.7%; Score 66; DB 1; Length 39;
 Best Local Similarity 73.3%; Pred. No. 0; Mismatches 4; Indels 0; Gaps 0;

Matches 11; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 GVKGDKGNPGPGAP 15
 Db 25 GVGXGXPGPAGP 39

RESULT 12

US-08-494-168-9

Sequence 4, Application US/08931820
 Patent No. 6010863
 GENERAL INFORMATION:

APPLICANT: Assay for collagen degradation

TITLE OF INVENTION: Assay for collagen degradation

NUMBER OF SEQUENCES: 4

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/931,820
 FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:
 APPLICATION NUMBER: EP 96202596.1

APPLICATION NUMBER: EP 96202596.1

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 1057 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: NO

ORIGINAL SOURCE:
 ORGANISM: Homo sapiens

TISSUE TYPE: Collagen type III

FEATURE:

NAME/KEY: Modified-site

LOCATION: 1055
; OTHER INFORMATION: /label= "Modified
; OTHER INFORMATION: /note= "Ala may be Pro"
; US-08-931-820-4

Query Match: Best Local Similarity 71.4%; Score 60; DB 3; Length 1057;
; Matches 10; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
; QY 1 GVKDKGNGWPGA 14
; ||||:|||||
; Db 710 GVKEERGSQGPGA 723

RESULT 14
US-08-903-825-21
; Sequence 21, Application US/08963825
; Patent No. 6110689

GENERAL INFORMATION:
; APPLICANT: Qvist, Per
; COMPUTER: Bonde, Martin
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments
; INVENTION: In Body Fluids, A Test Kit and Means for Carrying Out the
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
; TITLE OF INVENTION: Disorders Associated with the Metabolism of
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby & Darby PC
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022

COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/963,825
; FILING DATE:
; CLASSIFICATION:
; CLASSIFICATION:
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: 08/187,319
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Gogoris, Adda C
; REGISTRATION NUMBER: 29,714
; REFERENCE/DOCKET NUMBER: 4305/08701
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-527-7700
; TELEFAX: 212-753-6237
; TELEX: 236687

INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1078 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: COLLAGEN ALPHA 1 (III)
; US-09-570-573-21

Query Match: Best Local Similarity 71.4%; Score 60; DB 4; Length 1078;
; Matches 10; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
; QY 1 GVKDKGNGWPGA 14
; ||||:|||||
; Db 711 GVKEERGSQGPGA 724

RESULT 15
US-09-570-573-21
; Sequence 21, Application US/09570573
; Patent No. 6342561

GENERAL INFORMATION:
; APPLICANT: Qvist, Per
; COMPUTER: Bonde, Martin
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments
; INVENTION: In Body Fluids, A Test Kit and Means for Carrying Out the
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
; TITLE OF INVENTION: Disorders Associated with the Metabolism of
; NUMBER OF SEQUENCES: 21

INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1078 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: COLLAGEN ALPHA 1 (III)

Query Match: Best Local Similarity 71.4%; Score 60; DB 3; Length 1078;
; Matches 10; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
; QY 1 GVKDKGNGWPGA 14
; ||||:|||||
; Db 711 GVKEERGSQGPGA 724

RESULT 16
US-09-548-608-21
; Sequence 21, Application US/09548608
; Patent No. 635542

GENERAL INFORMATION:
; APPLICANT: Qvist, Per
; COMPUTER: Bonde, Martin
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments
; INVENTION: In Body Fluids, A Test Kit and Means for Carrying Out the
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
; TITLE OF INVENTION: Disorders Associated with the Metabolism of
; NUMBER OF SEQUENCES: 21

CORRESPONDENCE ADDRESS:

ADDRESSEE: Darby & Darby PC
 STREET: 805 Third Avenue
 CITY: New York
 STATE: New York
 COUNTRY: USA

ZIP: 10022

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/548,608

FILING DATE:

CLASSIFICATION:

PRIORITY APPLICATION DATA:

APPLICATION NUMBER: 08/187,319

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Goporis, Adda C

REGISTRATION NUMBER: 29,714

REFERENCE/DOCKET NUMBER: 4305/08701

INFORMATION FOR SEQ ID NO: 21:

SEQUENCE CHARACTERISTICS:

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

ORIGINAL SOURCE:

ORGANISM: Homo sapiens

IMMEDIATE SOURCE:

CLONE: COLLAGEN ALPHA 1 (III)

US-09-548-608-21

Query Match 65.2%; Score 60; DB 4; Length 1078;

Best Local Similarity 71.4%; Pred. No. 0.69; Matches 10; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GVKGDKGNGPGWCGA 14
Db 711 GVKGERGSPGGFGA 724

RESULT 17

US-08-359-889-24

Sequence 24, Application US/08359889B

Patient No. 597120

GENERAL INFORMATION:

APPLICANT: Reeders, Stephen T

APPLICANT: Morrison, Karen E

APPLICANT: Hudson, Billy G

TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides

FILE REFERENCE: 951263A

CURRENT APPLICATION NUMBER: US/08/359,889

CURRENT FILING DATE: 1995-03-07

EARLIER APPLICATION NUMBER: 07/621091

EARLIER FILING DATE: 1990-11-30

NUMBER OF SEQ ID NOS: 25

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 24

LENGTH: 471

TYPE: PRT

ORGANISM: Calf

US-08-359-889-24

Qy 1 GVKGDKGNGPGWCGA 15
Db 116 GMKGKGNNSGFPGPP 130

RESULT 18

US-09-167-364-24

Sequence 24, Application US/09167364

Patient No. 6007980

GENERAL INFORMATION:

APPLICANT: Reeders, Stephen T

APPLICANT: Morrison, Karen E

TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides

FILE REFERENCE: 951263B

CURRENT APPLICATION NUMBER: US/09/167,364

CURRENT FILING DATE: 1998-10-07

EARLIER APPLICATION NUMBER: 08/399889

EARLIER FILING DATE: 1995-03-07

NUMBER OF SEQ ID NOS: 25

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 24

LENGTH: 471

TYPE: PRT

ORGANISM: Calf

US-09-167-364-24

Qy 1 GVKGDKGNGPGWCGA 15
Db 116 GMKGKGNNSGFPGPP 130

RESULT 19

US-09-359-897-2

Sequence 2, Application US/09439897

Patient No. 6277558

GENERAL INFORMATION:

APPLICANT: Hudson, Billy G

TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides

FILE REFERENCE: 95-1263-C

CURRENT APPLICATION NUMBER: US/09/439,897

CURRENT FILING DATE: 1999-11-12

NUMBER OF SEQ ID NOS: 65

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 2

LENGTH: 471

TYPE: PRT

ORGANISM: Bos taurus

US-09-359-897-2

Qy 1 GVKGDKGNGPGWCGA 15
Db 116 GMKGKGNNSGFPGPP 130

RESULT 20

US-08-194-795-7

Sequence 7, Application US/08794795

Patient No. 5916766

GENERAL INFORMATION:

APPLICANT: Elshourbagy, Nabil

APPLICANT: Adamou, John

APPLICANT: Gross, Mitchell

APPLICANT: Lysko, Paul

US-08-194-795-7

Qy 1 GVKGDKGNGPGWCGA 15
Db 116 GMKGKGNNSGFPGPP 130

TITLE OF INVENTION: Human Macro Scavenger Rec
 TITLE OF INVENTION: eptor
 NUMBER OF SEQUENCES: 9
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: SmithKline Beecham Corporation
 STREET: 709 Sweetbriar Road
 CITY: King of Prussia
 STATE: PA
 COUNTRY: USA
 ZIP: 19406
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: DOS
 SOFTWARE: FastSEQ for Windows Version 2.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/794,795
 FILING DATE: 04-FEB-1997
 CLASSIFICATION: 435
 PRIORITY APPLICATION DATA:
 APPLICATION NUMBER: ATG50009P
 FILING DATE: 22-MAY-1996
 ATTORNEY/AGENT INFORMATION:
 NAME: Han, William T
 REGISTRATION NUMBER: 34,344
 REFERENCE/DOCKET NUMBER: ATG50009
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 610-270-5219
 TELEX: 610-270-4026
 INFORMATION FOR SEQ ID NO: 7:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 489 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 ; US-08-794-795-7

Query Match 60.9%; Score 56; DB 2; Length 489;
 Best Local Similarity 66.7%; Pred. No. 1.1; Mismatches 10; Conservative 1; Indels 4; Gaps 0; Gaps 0;
 Matches 10; Mismatches 1; Indels 4; Gaps 0;

Qy 1 GVKGDKGNGWPGAP 15
 |||||:|||_|
 Db 265 GVKGDQGKPGVQGV 279
 |||||:|||_|

RESULT 22-367B-2
 US-08-392-367B-2
 ; Sequence 2, Application US/08392367B
 ; Patent No. 5691197
 ; GENERAL INFORMATION:
 ; APPLICANT: Tryggvason, Karl
 ; APPLICANT: Elomaa, Outi
 ; APPLICANT: Kangas, Maari
 ; TITLE OF INVENTION: An Insolated DNA Sequence For a
 ; Patent No. 5691197
 ; TITLE OF INVENTION: No. 5691197e1 Macrophage Receptor with
 ; TITLE OF INVENTION: a Collagenous Domain and the
 ; TITLE OF INVENTION: Polypeptide Chain Encoded By
 ; TITLE OF INVENTION: such a Sequence
 ; NUMBER OF SEQUENCES: 2
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Fay, Sharpe, Beall, Fagan,
 ; ADDRESS: Minnich & McKee
 ; STREET: 1100 Superior Avenue
 ; STREET: Suite 700
 ; CITY: Cleveland
 ; STATE: Ohio
 ; COUNTRY: U.S.A.
 ; ZIP: 44114-2518
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette, 3.50 inch,
 MEDIUM TYPE: 720 Kb storable
 COMPUTER: IBM PS/2, Model 35 SX
 OPERATING SYSTEM: DOS 5.0
 SOFTWARE: Word Perfect 5.1
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/392,367B
 FILING DATE:
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Minnich, Richard J.
 REGISTRATION NUMBER: 24,175
 REFERENCE/DOCKET NUMBER: TRV 2 009
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (216) 861-5582

RESULT 21
 US-08-249-200-7
 ; Sequence 7, Application US/09249200
 ; Patent No. 6197931
 ; GENERAL INFORMATION:
 ; APPLICANT: EL-SHOURBAGY, NABIL
 ; APPLICANT: ADAMOU, JOHN
 ; APPLICANT: GROSS, MITCHELL
 ; APPLICANT: LYSK, PAUL
 ; TITLE OF INVENTION: HUMAN MARCO SCAVENGER RECEPTOR
 ; NUMBER OF SEQUENCES: 9
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESS: Ratner & Prestia
 ; STREET: P.O. Box 980
 ; CITY: Valley Forge
 ; STATE: PA
 ; COUNTRY: USA
 ; ZIP: 19482
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: DOS
 SOFTWARE: FastSEQ for Windows Version 2.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/249,200

TELEFAX: (216) 241-1666
 TELEX: (216) 980162
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 518 amino acids

TYPE: amino acid
 STRANDEDNESS: Single
 TOPOLOGY: Linear

; US-08-392-367B-2

Query Match, Local Similarity 60.9%; Score 56; DB 1; Length 518;
 Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 GVKGDKGNGPGWPGAP 15
 Db 294 GVKGDQKPGVQGVP 308

RESULT 23

; US-08-893-467A-2
 Sequence 2, Application US/08893467A

; PATENT NO. 6063901
 GENERAL INFORMATION:

; APPLICANT: Tryggason, Karl
 APPLICANT: Blomaa, Outi
 APPLICANT: Kangas, Marit

; TITLE OF INVENTION: An Insolated DNA Sequence For a
 Patent No. 6063901

; TITLE OF INVENTION: NO. 6063901el Macrophage Receptor with
 a Collagenous Domain and the
 Title of Invention: Polypeptide Chain Encoded by

; NUMBER OF SEQUENCES: 2
 CORRESPONDENCE ADDRESS:

; ADDRESSEE: Fay, Sharpe, Beall, Fagan,
 STREET: 1100 Superior Avenue
 CITY: Cleveland
 STATE: Ohio
 COUNTRY: U.S.A.

; ZIP: 44114-2518
 COMPUTER READABLE FORM:
 COMPUTER: IBM PS/2, Model 35 SX
 OPERATING SYSTEM: DOS 5.0
 SOFTWARE: Word Perfect 5.1
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/893, 467A
 FILING DATE:
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Minich, Richard J.
 REGISTRATION NUMBER: 28,665
 REFERENCE/DOCKET NUMBER: 40397/104/BABR
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (202)672-5300
 TELEFAX: (202)672-5399
 TELEX: 904136

; INFORMATION FOR SEQ ID NO: 10:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 546 amino acids
 TOPOLOGY: linear

; US-08-494-168-10

Query Match, Local Similarity 75.0%; Score 56; DB 1; Length 546;
 Matches 9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 4 GDKGNPGHGGAP 15
 Db 168 GPKGNPGVFGPP 179

RESULT 24

; US-08-494-168-10

; Sequence 10, Application US/08494168

; PATENT NO. 5731192
 GENERAL INFORMATION:

; APPLICANT: Reenders, Stephen T.

; TITLE OF INVENTION: Zhou, Jing

; TITLE OF INVENTION: Gene, Protein and Method

; NUMBER OF SEQUENCES: 10

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Foley & Lardner

; STREET: 300 K Street, N.W., Suite 500

; CITY: Washington, D.C.

; COUNTRY: USA

; ZIP: 20007-5109

; COMPUTER READABLE FORM:

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patientin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/494,168

; FILING DATE: 27-AUG-1993

; ATTORNEY/AGENT INFORMATION:

; NAME: Saxe, Bernhard D.

; REGISTRATION NUMBER: 28,665

; PRIORITY APPLICATION DATA:

; APPLICATION NUMBER: US 08/112,465

; REFERENCE/DOCKET NUMBER: 40397/104/BABR

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (202)672-5300

; TELEFAX: (202)672-5399

; TELEX: 904136

; INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:
 LENGTH: 546 amino acids

; TOPOLOGY: linear

; US-08-494-168-10

Query Match, Local Similarity 60.9%; Score 56; DB 1; Length 518;
 Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 GVKGDKGNGPGWPGAP 15
 Db 294 GVKGDQKPGVQGVP 308

RESULT 25

; US-09-011-735-2

; Sequence 2, Application US/09011735B

; PATENT NO. 6110708
 GENERAL INFORMATION:

; APPLICANT: Wakamatsu, No. 6110708utaka

; TITLE OF INVENTION: Recombinant Conglutinin and Producing Method Thereof

; FILE REFERENCE: 19056/30548

; CURRENT APPLICATION NUMBER: US/09/011-735B

; CURRENT FILING DATE: 1998-05-22

; EARLIER FILING DATE: 1995-08-17

; NUMBER OF SEQ ID NOS: 6

; SOFTWARE: Patentin Ver. 2.0

; SEQ ID NO: 2

; LENGTH: 171

; US 08-893-467A-2

Query Match, Local Similarity 60.9%; Score 56; DB 3; Length 518;
 Best Local Similarity 66.7%; Pred. No. 1.2; Mismatches 4; Indels 0; Gaps 0;
 Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

TYPE: PRT
; ORGANISM: Bovine
; US-09-011-735-2

Query Match 59.8%; Score 55; DB 3; Length 171;
; Best Local Similarity 66.7%; Pred. No. 0.53;
; Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
; Qy 1 GVKGDKGNPGWPGAP 15
; Db 112 GLKGEKGAGGETGAP 126

RESULT 26
US-09-011-735-1
; Sequence 1, Application US/09011735B
; Patent No. 6110708
; GENERAL INFORMATION:
; APPLICANT: Wakamiya, NO. 6110708utaka
; TITLE OF INVENTION: Recombinant Conglutinin and Producing Method Thereof
; FILE NUMBER: 19036/34548
; CURRENT APPLICATION NUMBER: US/09/011,735B
; EARLIER FILING DATE: 1995-08-17
; NUMBER OF SEQ ID NOS: 6
; SEQ ID NO 1 SOFTWARE: PatentIn Ver. 2.0
; LENGTH: 351
; TYPE: PRT
; ORGANISM: Bovine
; OS-09-011-735-1

RESULT 27
US-08-494-168-2
; Sequence 2, Application US/08494168
; Patent No. 5731192
; GENERAL INFORMATION:
; APPLICANT: Reeders, Stephen T.
; APPLICANT: Zhou, Jing
; TITLE OF INVENTION: Detecting Collagen Deficiency
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER: IBM PC compatible
; COMPUTER TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/494,168
; FILING DATE:
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/112,465
; FILING DATE: 27-AUG-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Saxe, Bernard D.
; REGISTRATION NUMBER: 28 665

Query Match 59.8%; Score 55; DB 3; Length 171;
; Best Local Similarity 66.7%; Pred. No. 0.53;
; Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
; Qy 1 GVKGDKGNPGWPGAP 15
; Db 112 GLKGEKGAGGETGAP 126

RESULT 28
US-09-219-849-38
; Sequence 38, Application US/09219849
; Patent No. 6150081
; GENERAL INFORMATION:
; APPLICANT: VAN HEERDE, GEORGE V.
; APPLICANT: VAN RIJN, ALEXIS C.
; APPLICANT: BOUTSTRA, JAN B.
; APPLICANT: DE WOLF, FREDERIK A.
; APPLICANT: MOOBROEK, ANDREAS
; APPLICANT: WERTEN, MARC W.T.
; APPLICANT: WIND, RICHELE D.
; APPLICANT: VAN DEN BOSCH, TANJA J.
; TITLE OF INVENTION: SILVER HALIDE EMULSIONS WITH RECOMBINANT COLLAGEN
; TITLE OF INVENTION: SUITABLE FOR PHOTOGRAPHIC APPLICATION AND ALSO THE
; FILE REFERENCE: 2728-2
; CURRENT APPLICATION NUMBER: US/09/219,849
; CURRENT FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 38
; LENGTH: 228
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence
; ; OTHER INFORMATION: amino acid sequence
; US-09-219-849-38

RESULT 29
US-08-794-795-2
; Sequence 2, Application US/08794795
; Patent No. 5916766
; GENERAL INFORMATION:
; APPLICANT: Elshourbagy, Nabil
; APPLICANT: Adamou, John
; APPLICANT: Gross, Mitchell
; APPLICANT: Lysko, Paul
; TITLE OF INVENTION: Human Macro Scavenger Rec
; TITLE OF INVENTION: eptor
; NUMBER OF SEQUENCES: 9

REFERENCE/DOCKET NUMBER: 40397/104/BABR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 672-5300
; TELEFAX: (202) 672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1694 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-494-168-2

CORRESPONDENCE ADDRESS:
 ADDRESSEE: SmithKline Beecham Corporation
 STREET: 709 Swede-hill Road
 CITY: King of Prussia
 STATE: PA
 COUNTRY: USA
 ZIP: 19406
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: DOS
 SOFTWARE: FASTSEQ for Windows version 2.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/794,795
 FILING DATE: 04-FEB-1997
 CLASSIFICATION: 435
 PRIORITY APPLICATION DATA:
 APPLICATION NUMBER: ATG50009P
 FILING DATE: 22-MAY-1996
 ATTORNEY/AGENT INFORMATION:
 NAME: Han, William T
 REGISTRATION NUMBER: 34,344
 REFERENCE/DOCKET NUMBER: ATG50009
 TELEPHONE: 610-270-5219
 TELEX: 610-270-4026
 INFORMATION FOR SEQ ID NO: 2:
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 495 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 ; US-09-249-200-2
 Query Match 57.6%; Score 53; DB 2; Length 495;
 Best Local Similarity 60.0%; Pred. No. 3.1;
 Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;
 QY 1 GVKGDKGNGPGWPGAP 15
 | | | : | | |
 Db 270 GAKGDQGQPGLOGVP 284
 | | | : | | |
 ; RESULT 30
 ; US-09-249-200-2
 ; Sequence 2, Application US/09249200
 ; Patient No. 6197931
 ; GENERAL INFORMATION:
 ; APPLICANT: ELSHOURBACY, NABIL
 ; APPLICANT: ADAMOU, JOHN
 ; APPLICANT: GROSS, MITCHELL
 ; APPLICANT: LYSKO, PAUL
 ; TITLE OF INVENTION: HUMAN MARCO SCAVENGER RECEPTOR
 ; NUMBER OF SEQUENCES: 9
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Rarner & Prestia
 ; STREET: P.O. Box 980
 ; CITY: Valley Forge
 ; STATE: PA
 ; COUNTRY: USA
 ; ZIP: 19482
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: DOS
 ; SOFTWARE: FASTSEQ for Windows Version 2.0
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/249,200
 ; FILING DATE: 12-FEB-1999
 ; CLASSIFICATION:
 ; PRIORITY APPLICATION DATA:

APPLICATION NUMBER: 08/794,795
 FILING DATE: 04-FEB-1997
 APPLICATION NUMBER: 60/017,699
 FILING DATE: 23-MAY-1996
 ATTORNEY/AGENT INFORMATION:
 NAME: Prestia, Paul F
 REGISTRATION NUMBER: 23,031
 REFERENCE/DOCKET NUMBER: ATG-50009-1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 610-407-0700
 TELEX: 846169
 INFORMATION FOR SEQ ID NO: 2:
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 495 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 ; US-09-249-200-2
 Query Match 57.6%; Score 53; DB 4; Length 495;
 Best Local Similarity 60.0%; Pred. No. 3.1;
 Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;
 QY 1 GVKGDKGNGPGWPGAP 15
 | | | : | | |
 Db 270 GAKGDQGQPGLOGVP 284
 | | | : | | |
 Search completed: November 1, 2002, 12:54:44
 Job time : 11.5 secs

PAGE BLANK (USP0)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

Run on: November 1, 2002, 12:38:32 ; Search time 13.5 Seconds
Title: OM protein - protein search, using sw model.

Post-processing: Minimum Match 0%
Maximum DB seq length: 0
Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Title: Perfect score: 92
Sequence: 1 GVKGDKGNPGWPGAP 15

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : PIR 71:
1: pir1:
2: pir2:
3: pir3:
4: pir4:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No. Score Query Match Length DB ID Description

RESULT 1
CGII04B
collagen alpha 1(IV) chain precursor - human
N:Alternate names: procollagen alpha 1(IV) chain
C:Species: Homo sapiens (man)
C:Date: 28-May-1986 #sequence_revision 31-Dec-1992 #text_change 07-Dec-1999
C:Accession: Si6876; A32117; S02738; S00048; S25826; A23115; S00207; S39614; A02863;
R:Soininen, R.; Huotari, M.; Ganguly, A.; Prockop, D.J.; Tryggvason, K.
J. Biol. Chem. 264, 13565-13571, 1989
A:Title: Structural organization of the gene for the alpha-1 chain of human type IV collagen
A:Reference number: S16876; MUID:89340433
A:Accession: S16876
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1659 <S011>
A:Cross-references: EMBL:J04217; GB:J05039; NID:9180800; PIDN:AAA53098_1; PID:9180803
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1988
R:Soininen, R.; Huotari, M.; Hostikka, S.L.; Prockop, D.J.; Tryggvason, K.
J. Biol. Chem. 263, 17217-17220, 1988
A:Title: The structural genes for alpha1 and alpha2 chains of human type IV collagen
A:Reference number: A92690; MUID:8934231
A:Accession: A32117
A:Molecule type: DNA
A:Residues: 1-28 <S012>
A:Cross-references: EMBL:J04217; NID:9180759; PIDN:AAA53097_1; PID:9533233
R:Poeschl, E.; Pollner, R.; Kuehn, K.
EMBO J. 7, 2687-2695, 1988
A:Title: The genes for the alpha1(IV) and alpha2(IV) chains of human basement membrane
A:Reference number: S02738; MUID:89030632
A:Accession: S02738
A:Status: translation not shown
A:Molecule type: DNA
A:Residues: 1-6, 'L', 8-28 <POE>
A:Cross-references: EMBL:X12784; NID:g30072
R:Brazel, D.; Oberbaumer, I.; Dierlinger, H.; Babel, W.; Glanville, R.W.; Deutzhmann, P.
Eur. J. Biochem. 168, 529-536, 1987
A:Title: Completion of the amino acid sequence of the alpha1 chain of human basement membrane
A:Reference number: S00048; MUID:88029471
A:Accession: S00048
A:Molecule type: mRNA
A:Residues: 1-318; 'A', 320-944 <BRAl>
A:Cross-references: EMBL:X05561; NID:g30066; PIDN:CAA29075_1; PID:930067
A:Accession: S25826
A:Molecule type: protein
A:Residues: 271-318; 'A', 320-554 <BRAl>
R:Glanville, R.W.; Olan, R.O.; Siebold, B.; Ristelli, J.; Kuehn, K.
Eur. J. Biochem. 152, 213-219, 1985
A:Title: Amino acid sequence of the N-terminal aggregation and cross-linking region
A:Reference number: A23115; MUID:86004708
A:Accession: A23115
A:Molecule type: protein
A:Residues: 28-236, 'KE', 239-240, 'K', 242-243 <GLA>

collagen alpha 1(IV) chain precursor - human
type VII collagen
hypothetical protein
protein F5633_1
collagen 1 - marin macrophage bacteriophage
collagen alpha 2(I) collagen alpha 1(X)
macrophage scavenging macrophage scaveng

ALIGNMENTS

30	58	63.0	285	2	S60598
31	58	63.0	464	2	S59513
32	58	63.0	1758	2	T29350
33	58	63.0	1759	2	T29351
34	57	62.0	298	2	JC1448
35	57	62.0	650	2	T22002
36	57	62.0	920	2	A45748
37	57	62.0	1349	2	I48103
38	56	60.9	299	2	T28333
39	56	60.9	371	2	E8633
40	56	60.9	380	2	A48295
41	56	60.9	518	2	A52840
42	56	60.9	775	2	A61228
43	56	60.9	1142	2	JX03369
44	55	60.3	350	2	I73338
45	55.5	60.3	458	2	B44407

A; Experimental source: placenta
A; Note: the amino end of the mature form is blocked
C; Complex: type IV collagen is a heterotrimer of two alpha 1(IV) chains and one alpha 2(IV) chain.
R; Sczinian, R.; Hata-Risku, T.; Prockop, D.J.; Tryggrason, K.
FEBS Lett. 225: 188-194, 1987.

A; Title: Complete primary structure of the alpha(1)-chain of human basement membrane (type IV)
A; Reference number: S00207; MUID:88083584

A; Accession: S00207
A; Molecule type: mRNA

A; Residues: 244-530 <S013>

A; Cross-references: EMBL:Y00706; NID:929548; PIDN:CAA68698.1; PID:929549

A; Accession: A02863; MUID:85003629

A; Molecule type: protein

A; Residues: 554-718; D720-836, 'Y', 838-841, 'P', 843-903, 'Q', 905-913, 'K', 915-997, 'K', 999-
A; Experimental source: Placenta
R; Gianville, R.W.; Rauter, A.
Hoppe-Seyler's Z. Physiol. Chem. 362, 943-951, 1981

A; Title: Pepsin fragments of human basement-membrane collagens showing interruptions
A; Reference number: S16508; MUID:82005835

A; Accession: A58517

A; Molecule type: protein

A; Residues: 554-537, 'G', 539, 'G', 541-542, 'X', 544-553; 1389-1405, 'X', 1408-1409, 'X', 1411-1412, 1661-Domains: collagen IV carboxyl-terminal nonhelical repeat <CTP2>
F; 1-26/Modified site: blocked amino end (Ala) (in mature form) #status experimental

R; McWright, R.S.; Benson, V.A.; Lovello, K.T.; van der Rest, M.; Fietzek, P.P.
Biochemistry 22, 4340-4348, 1983

A; Title: Isolation and characterization of pepsin-solubilized human basement membrane (type IV)
A; Reference number: S16910; MUID:84053346

A; Molecule type: protein

A; Residues: 554-537, 'G', 539, 'G', 541-542, 'G', 544-549; 939-940, 'M', 942-944, 'V', 946, 'X', 948-
A; Experimental source: Placenta
R; Pihalainen, T.; Trygvason, K.; Myers, J.C.; Kurkinen, M.; Lebo, R.; Cheung, M.C.; Eng, J. Biol. Chem. 260, 7681-7687, 1985

A; Title: cDNA clones coding for the Pro-alpha-1(IV) chain of human type IV procollagen
A; Reference number: S01466; MUID:85207819

A; Accession: S01466
A; Molecule type: mRNA

A; Residues: 1556-1669 <PIH>

A; Cross-references: EMBL:M10940; NID:9180421; PIDN:AAA52006.1; PID:9180424
R; Brinker, J.M.; Gudas, L.J.; Loesl, H.R.; Wang, S.Y.; Rosenblloom, J.; Kefalides, N.A.; Proc. Natl. Acad. Sci. U.S.A. 82, 3645-3653, 1985

A; Title: Restricted homology between human alpha-1 type IV and other procollagen chains.
A; Reference number: S16879; MUID:85216555

A; Accession: S16879
A; Molecule type: mRNA

A; Residues: 1559-1669 <BRI>

A; Cross-references: EMBL:M11315; NID:9180817; PIDN:AAA52042.1; PID:9180818
R; Oberbaumer, I.; Laurent, M.; Schwarz, U.; Sakurai, Y.; Yamada, Y.; Vogeli, G.; Voss, Eur. J. Biochem. 147, 217-224, 1985

A; Title: Amino acid sequence of the non-collagenous globular domain (NC1) of the alpha-1 chain of the basement membrane (type IV)
A; Reference number: A02664; MUID:85127033

A; Accession: S19091
A; Molecule type: protein

A; Residues: 1435-1461, 'H', 1463-1482, 'X', 1484-1491; 1501-1514, 'X', 1516-1519; 1534-1553, 'X',
R; Siebold, B.; Deutermann, R.; Kuehn, K.
Eur. J. Biochem. 176, 617-624, 1988

A; Title: The arrangement of intra and intermolecular disulfide bonds in the carboxyterminal
A; Reference number: S02450; MUID:89005112

A; Cross-references: annotation; disulfide bonds
C; Genetics: GDB:COL4A1

A; Cross-references: GDB:119791; OMIM:120130
A; Map position: 13q4-13q34

A; Introns: 28/3; 48/3; 78/3; 93/3; 108/3; 129/3; 147/3; 156/3; 184/3; 205/3; 217/3; 231/

RESULT 2

CGMSAB

collagen alpha 1(IV) chain precursor - mouse

C; Species: Mus musculus (house mouse)

C; Date: 28-May-1986 #sequence_revision 31-Dec-1992 #text_change 16-Jun-2000

C; Accession: A33525; S01454; A28066; A02864; A52536; A29301; S19079; A32003; A31766; R; Muthukumaran, G.; Blumberg, B.; Kurkinen, M.
J. Biol. Chem. 264, 6310-6311, 1989

A; Title: The complete primary structure for the alpha-1-chain of mouse collagen IV. D
A; Reference number: A33525; MUID:89197932

A; Accession: A33525
A; Molecule type: mRNA

A; Residues: 1-1669 <NUT>

A; Cross-references: EMBL:J06694; NID:556296; PIDN:AAA50292.1; PID:9556297
R; Wood, L.; Theriault, N.; Vogeli, G.
FEBS Lett. 227, 5-8, 1988

A; Title: cDNA clones completing the nucleotide and derived amino acid sequence of the
A; Reference number: S01454; MUID:88112221

A; Accession: S01454
A; Molecule type: mRNA

A;Residues: 1-185, 'L', 187-318, 'S', 320-368, 'L', 370-402, 'F', 404-480, 'L', 482-492, 'H', 494-71
A;Cross-references: EMBL:06777
R;Killen, P.D.; Burbelo, P.; Sakurai, Y.; Yamada, Y.
J. Biol. Chem. 263, 8706-8709, 1988

A;Title: structure of the amino-terminal portion of the murine alpha-1(IV) collagen chain
A;Reference number: A28066; MUID:88243724
A;Accession: A28066
A;Molecule type: mRNA
A;Residues: 1-129 <K1>
A;Cross-references: EMBL:J03758; NID:9192669; PIDN:AAA37439.1; PID:9192670
R;Oberbaumer, I.; Laurent, M.; Schwarz, U.; Sakurai, Y.; Yamada, Y.; Vogeli, G.; Voss,
J. Biol. Chem. 147, 217-224, 1985

A;Title: Amino acid sequence of the non-collagenous globular domain (NC1) of the alpha-1
A;Reference number: A02864; MUID:85127033
A;Accession: A02864
A;Molecule type: mRNA
A;Residues: 1276-1669 <OBE>
A;Cross-references: EMBL:X02201; NID:950233; PIDN:CAA26132.1; PID:91333876
R;Nath, P.; Laurent, M.; Horn, E.; Sobel, M.E.; Zon, G.; Vogeli, G.
Gene 43, 301-304, 1986

A;Title: Isolation of an alpha-1 type-IV collagen cDNA clone using a synthetic oligodeox
A;Reference number: A25636; MUID:86501886
A;Accession: A25636
A;Molecule type: mRNA
A;Residues: 1149-1396, 'S', 1398-1424 <NAT>
A;Cross-references: EMBL:MA4042; NID:9192286; PIDN:AAA37342.1; PID:9192287
A;Note: the authors translated the codon CAG for residue 1374 as Arg
R;Kurkinen, M.; Condon, M.R.; Blumberg, B.; Barlow, D.P.; Quinones, S.; Saus, J.; Pihlaj
J. Biol. Chem. 262, 8496-8499, 1987

A;Title: Extensive homology between the carboxyl-terminal peptides of mouse alpha-1(IV)
A;Reference number: A94680; MUID:87250460
A;Accession: A23501
A;Molecule type: mRNA
A;Cross-references: EMBL:MI5832; NID:9192282; PIDN:AAA37340.1; PID:9387115
R;Killen, P.D.; Burbelo, P.; Martin, G.R.; Yamada, Y.
J. Biol. Chem. 263, 12310-12314, 1988

A;Title: Characterization of the promoter for the alpha-1(IV) collagen gene. DNA sequenc
A;Reference number: S19079; MUID:88315019
A;Accession: S19079
A;Molecule type: DNA
A;Residues: 1-28 <K1>
A;Cross-references: EMBL:J03944; NID:9192673; PIDN:AAA7442.1; PID:9466503
R;Kaytes, P.; Wood, L.; Theriault, N.; Kurkinen, M.; Vogeli, G.
J. Biol. Chem. 263, 19274-19277, 1988

A;Title: Head-to-head arrangement of murine type IV collagen genes.
A;Reference number: A92702; MUID:89066738
A;Accession: A32003
A;Molecule type: DNA
A;Residues: 1-28 <KAY>
A;Cross-references: EMBL:J04448; NID:9192666; PIDN:AAA37437.1; PID:9450449
R;Burbelo, P.D.; Martin, G.R.; Yamada, Y.
Proc. Natl. Acad. Sci. U.S.A. 85, 9679-9682, 1988

A;Title: Alpha1(IV) and alpha2(IV) collagen genes are regulated by a bidirectional prom
A;Reference number: A94220; MUID:89071759
A;Accession: A31766
A;Molecule type: DNA
A;Residues: 1-28 <BUR>
A;Cross-references: EMBL:M23333; NID:9340878; PIDN:AAA51625.1; PID:9535668
R;Sakurai, Y.; Sullivan, M.; Yamada, Y.
J. Biol. Chem. 261, 6654-6657, 1986

A;Title: Alpha-1 type IV collagen gene evolved differently from fibrillar collagen genes
A;Reference number: S19094; MUID:86196099
A;Molecule type: DNA
A;Residues: 110-1135,1189-1316,1342-1383,1418-1487 <SAK>
A;Cross-references: EMBL:M13027
R;Schuppan, D.; Timpl, R.; Glanville, R.W.
FEBS Lett. 115, 297-300, 1980

A;Title: Discontinuities in the triple helical sequence Gly-X-Y of basement membrane (ty
A;Reference number: S16909; MUID:80246483
A;Accession: S16909
A;Molecule type: protein

A;Residues: 940-946, 'G', 948-949, 'G', 951-955, 'G', 957-1213-1228, 'X', 1230-1234, 'P', 1236-
A;Cross-references: EMBL:R06777
R;Schuppan, D.; Glanville, R.W.; Timpl, R.
R;Killen, P.D.; Burbelo, P.; Sakurai, Y.; Yamada, Y.
R; Oberbaumer, I.; Laurent, M.; Schwarz, U.; Sakurai, Y.; Yamada, Y.; Vogeli, G.; Voss,
J. Biol. Chem. 263, 505-512, 1982

A;Title: Covalent structure of mouse type-IV collagen. Isolation, order and partial a
A;Reference number: A25991; MUID:82186723
A;Accession: A25991
A;Molecule type: protein
A;Residues: 940-946, 'X', 948-949, 'X', 951-955, 'X', 957-964, 'X', 966-991, 'X', 993-1003, 'X'
A;Accession: B25991
A;Molecule type: protein
A;Residues: 1173-1181, 'X', 1183-1184, 'X', 1185-1187, 'X', 1189-1205, 'O', 1207, 'X', 1210-12
3, 'SP', 1266, 'IT', 1269, 'SK', 1272, 'DM', 1275, 'L', 1277-1282, 1316-1318, 'X', 1320-1327, 'X', 1
R;Weber, S.; Engel, J.; Wiedemann, H.; Glanville, R.W.; Timpl, R.
R;J. Biol. Chem. 139, 401-410, 1984

A;Title: Subunit structure and assembly of the globular domain of basement-membrane c
A;Reference number: S17801; MUID:84132058
A;Accession: S17801
A;Molecule type: protein
A;Residues: 1435-1443 <WEB>
A;Genetics:
A;Introns: 28/3; 48/3; 78/3; 93/3; 108/3; 129/3
A;Note: the list of introns may be incomplete
C;Superfamily: collagen alpha 1(IV) chain
C;Keywords: basement membrane; cell binding; coiled coil; duplication; extracellular
F;1-27/Domain: signal sequence #status predicted <IG>
F;28-169/Product: collagen alpha 1(IV) chain #status predicted <NAT>
F;28-162/Domain: 7S <7SD>
F;163-1440/Domain: collagenous, triple helix <COL>
F;159-599/Region: cell attachment (R-G-D) motif
F;781-783/Region: cell attachment (R-G-D) motif
F;918-919/Region: cell attachment (R-G-D) motif
F;968-970/Region: cell attachment (R-G-D) motif
F;141-169/Domain: carboxyl-terminal nonhelical, NC1 <NC1>
F;1441-1552/Region: duplication
F;31-36, 39, 41, 434, 467, 470/Disulfide bonds: interchain #status predicted
F;126/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;971-974, 977, 986, 989, 1001, 1007, 1019, 1022, 1031, 1037, 1040, 1055, 1060, 1063, 1075, 1078, 109
92, 1298, 1310, 1313, 1322, 1337, 1346, 1349, 1422, 1425, 1431, 1437, 1440/Modified site: hydroxy
F;1214-1424/Modified site: 4-hydroxyproline (Pro) #status experimental
F;1304/Modified site: 5-hydroxylysine (Lys) #status experimental
F;1505-1511, 1616-1622/Disulfide bonds: #status predicted
F;3126/Binding site: carbonhydrate (Asn) (covalent) #status predicted
F;971-974, 977, 986, 989, 1001, 1007, 1019, 1022, 1031, 1037, 1040, 1055, 1060, 1063, 1075, 1078, 109
92, 1298, 1310, 1313, 1322, 1337, 1346, 1349, 1422, 1425, 1431, 1437, 1440/Modified site: hydroxy
F;1214-1424/Modified site: 4-hydroxyproline (Pro) #status experimental
F;1505-1511, 1616-1622/Disulfide bonds: #status predicted
Query Match Score 84.8%; Best Local Similarity 86.7%; Pred. No. 0; Mismatches 0; Indels 0; Gaps 0;
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 GVKGDKGNGPGPGAP 15
Db 1263 GPKGDKGNQPGAP 1277

RESULT 3
S16907
collagen alpha 1(IV) chain - bovine (fragments)
C;Species: Bos primigenius tauris (cattle)
C;Date: 04-Dec-1992 #sequence_revision 03-Apr-1995 #text_change 19-Oct-1995
C;Accession: A39474; S16907; S58432; A39419
R;Gunwar, S.; Neelken, M.E.; Hudson, B.G.
J. Biol. Chem. 266, 14088-14094, 1991
A;Title: Properties of the collagenous domain of the alpha3(IV) chain, the goodpastur
A;Reference number: A39474; MUID:91310700
A;Accession: A39474
A;Molecule type: protein
A;Residues: 1-9 <GU2>
R;Schuppan, D.; Glanville, R.W.; Timpl, R.; Dixit, S.N.; Kang, A.H.
R;Schuppan, D.; Glanville, R.W.; Timpl, R.; Dixit, S.N.; Kang, A.H.
R;Killen, P.D.; Burbelo, P.; Sakurai, Y.; Yamada, Y.
R; Oberbaumer, I.; Laurent, M.; Schwarz, U.; Sakurai, Y.; Yamada, Y.; Vogeli, G.; Voss,
J. Biol. Chem. 263, 505-512, 1982

A;Title: Sequence comparison of pepsin-resistant segments of basement-membrane colla
A;Reference number: S16907; MUID:84256630
A;Accession: S16907
A;Molecule type: protein

A;Residues: 4-8, 'E' 10-68;69-158;159-221;322-349 <SC2>
 R;Butkowski, R.J.; Langerveld, J.P.M.; Wieslander, J.; Hamilton, J.; Hudson, B.G.
 J. Biol. Chem. 262, 7874-7877, 1987

A;Title: Localization of the Goodpasture epitope to a novel chain of basement membrane
 A;Reference number: S18432; MUID:87222419

A;Accession: S18432

A;Molecule type: protein
 A;Residues: 337-347, 'FL', 350-353 <BU2>
 C;Superfamily: collagen alpha 1(IV) chain
 C;Keywords: basement membrane; cell binding; coiled coil; disulfide bond; extracellular
 F;31;34;37,46;61,69;78,84;87,102;110-122;125;137;140;143;149;155;158;161;164;185;188;199
 proline (Pro) #status experimental

Query Match	78 3%	Score 72;	DB 2;	Length 356;
Best Local Similarity	66 7%	Pred. No.	0.003;	
Matches	10;	Conservative	3;	Mismatches 2;
				Indels 0;
				Gaps 0;

OY 1 GVKDDKGNGWPGAP 15
 |:|||||:|||||||:
 Db 248 GLXGERGNPGWPGRP 262

RESULT 4

S22917

collagen alpha 5(IV) chain precursor, renal splice form - human

N;Alternate names: procollagen alpha 5(IV) chain

N;Contains: collagen alpha 5(IV) chain precursor, leukocyte splice form

C;Species: Homo sapiens (man)

C;Date: 30-Sep-1993 #sequence_revision 27-Feb-1997 #ext_change 21-Jul-2000

C;Accession: S22917; A34365; A57079; A37122; A53171; A34850; S18850; 156971; 176598; A35
 R; Zhou, J.; Hertz, J.M.; Leinonen, A.; Tryggvason, K.
 J. Biol. Chem. 267, 12475-12481, 1992

A;Title: Complete amino acid sequence of the human alpha-5(IV) collagen chain and identi
 n Alport syndrome patient

A;Reference number: S22917; MUID:92316923

A;Accession: S22917

A;Molecule type: mRNA

A;Residues: 1-967 <ZHO>

A;Cross-references: GB:M90464; NID:gi180826; PIDN:AAA52045.1; PID:gi180825

A;Title: Structure of the human type IV collagen COL4A5 gene.

A;Reference number: A54365; MUID:94165049

J. Biol. Chem. 269, 6608-6614, 1994

A;Title: Structure of the human type IV collagen COL4A5 gene.

A;Accession: A54365

A;Molecule type: DNA

A;Residues: 1-922 <ZH2>

A;Cross-references: GB:U04470; NID:gi463378; GB:U04520; NID:gi463428; PIDN:AAC27815.1; PID
 R; Zhou, J.; Mochizuki, T.; Smeets, H.; Antignac, C.; Laurila, P.; de Paepe, A.; Tryggvas
 Science 261, 1167-1169, 1993

A;Title: Deletion of the paired alpha5(IV) and alpha6(IV) collagen genes in inherited sm

A;Reference number: A57079; MUID:93361972

A;Accession: A57079

A;Molecule type: DNA

A;Residues: 1-27 <ZH4>

A;Cross-references: GB:Z37153; NID:gi587203; PIDN:CAA85512.1; PID:gi587204

A;Title: Complete primary structure of the triple-helical region and the carboxyl-termin

A;Reference number: A37122; MUID:90337990

A;Accession: A37122

A;Molecule type: mRNA

A;Residues: 8d-439, 'GS' 442-624, 'LALQ' 629-666, 'FR' 669-887, 'R' 889-1264, 1271-1691 <PIH>

A;Cross-references: GB:J05558; MBD:MB8526; NID:gi134209

A;Note: submitted to the EMBL Data Library, February 1991

A;Note: the authors translated the codon GCG for residue 115 as Val

R;Renieri, A.; Seri, M.; Myers, J.C.; Pihlajaniemi, T.; Massella, L.; Rizzoni, G.; De Ma

Hum. Mol. Genet. 1, 127-129, 1992

A;Title: De novo mutation in the COL4A5 gene converting glycine 325 to glutamic acid

A;Reference number: 154317; MUID:93244772

A;Accession: 154317

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: mRNA

A;Residues: 313-324, 'E' 326-330 <REN>

A;Cross-references: GB:S59334; NID:929946; PIDN:AD13909.1; PID:g4261609

R;Honstiika, S.I.; Boddy, R.L.; Byers, M.G.; Hoeynya, M.; Shows, T.B.; Tryggvason, K.
 Proc. Natl. Acad. Sci. U.S.A. 87, 1605-1610, 1990

A;Title: Identification of a distinct type IV collagen alpha chain with restricted k1

A;Reference number: A34850; MUID:90160375

A;Accession: A34850

A;Molecule type: mRNA

A;Residues: 914-1264,1271-1591 <HOS>

A;Cross-references: EMBL: M31115; NID:gi180824; PIDN:AAA52045.1; PID:gi180825

R;Honstiika, S.I.; Boddy, R.L.; Chow, L.T.; Tryggvason, K.
 Genomics 9, 1-9, 1991

A;Title: Characterization of the 3' half of the human type IV collagen alpha-5 gene t

A;Reference number: A37969; MUID:91169491

A;Accession: S18850

A;Molecule type: DNA

A;Residues: 924-1264, 1271-1691 <ZH3>

A;Cross-references: EMBL: M63470; EMBL: M63471; EMBL: M63472; EMBL: M63473; NID:gi17922; PIDN:AAA51558.1;
 8; EMBL: M63470; EMBL: M63471; EMBL: M63472; EMBL: M63473; NID:gi17922; PIDN:AAA51558.1;
 Guo, C.; Van Damme, B.; Van Den Damme-Lombaerts, R.; Van den Berghe, H.; Cassiman, J.J.
 Kidney Int. 44, 1316-1321, 1993

A;Title: Differential splicing of COL4A5 mRNA in kidney and white blood cells: a comp

A;Reference number: 156971; MUID:94133540

A;Accession: 156971

A;Status: translated from GB/EMBL/DDBJ

A;Molecule type: mRNA

A;Residues: 1258-176 <QU01>

A;Cross-references: GB:S69168; NID:gi545095; PIDN: AAC60612.1; PID:gi545096

A;Title: kidney splice form

A;Accession: 176598

A;Status: translated from GB/EMBL/DDBJ

A;Molecule type: mRNA

A;Residues: 1284-1291, 'TFLGIVIACLV' <GU02>

A;Cross-references: GB:S69169; NID:gi45097; PIDN: AAC60613.1; PID:gi545098

A;Title: frameshift mutation in patient with Alport syndrome

R;Myers, J.C.; Jones, T.A.; Pohjolainen, E.R.; Kadri, A.S.; Goddard, A.D.; Sheer, D.;
 Am. J. Hum. Genet. 46, 1024-1033, 1990

A;Title: Molecular cloning of alpha5(IV) collagen and assignment of the gene to the r

A;Accession: A35335

A;Status: nucleic acid sequence not shown

A;Molecule type: mRNA

A;Residues: 1448-1477 <AYE>

R; Nakazato, H.; Sato, S.; Ushijima, T.; Matsuuwa, T.; Koitabashi, Y.; Takada, T.;
 Kidney Int. 46, 1307-1314, 1994

A;Title: Mutations in the COL4A5 gene in Alport syndrome: a possible mutation in prim

A;Reference number: 156975; MUID:95156893

A;Accession: 156975

A;Status: translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1595-1602 <NAK>

A;Cross-References: GB:S775903; NID:gi913882; PIDN: AAB33374.1; PID:gi913883

A;Note: premature termination mutation from a patient with Alport syndrome; one other
 R; Lemmink, H.H.; Schroeder, C.H.; Brunner, H.G.; Nelen, M.R.; Zhou, J.; Tryggvason, K.
 Genetics 177, 485-489, 1993

A;Title: Identification of four novel mutations in the COL4A5 gene of patients with A

A;Reference number: 154188; MUID:94010948

A;Accession: 154188

A;Status: translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1604-1607, 'VHDARKC' <LEM>

A;Cross-references: GB:S65767; NID:gi25563; PIDN: AD13967.1; PID:g4261667

A;Note: frameshift mutation from a patient with Alport syndrome; five other mutations
 C;Comment: Prolines and Lysines at the third position of the tripeptide repeating uni
 ed and subsequently O-glycosylated.

C;Genetics:

A;Gene: GDB:COL4A5; AT5

A,Cross-references: GDB:120596; OMIM:303630
A,Map position: Xq22-Xq22
A,Introns: 27/3; 47/3; 77/3; 92/3; 107/3; 128/3; 146/3; 155/3; 182/3; 203/3; 215/3; 229/3; 79/1; 837/1; 923/1; 973/1; 1006/1; 1036/1; 1082/3; 1125/1; 1152/1; 1185/1; 1
C>Note: the alpha 5(IV) and alpha 6(IV) chain genes are encoded on opposite strands with
C:Complex: This minor type IV collagen is thought to form a heterotrimer of two alpha 5(IV)
among trimer amino-terminal domains (with disulfide and desmosine cross-links), dimeric a
ter associations in the interrupted helical domain (with disulfide and desmosine cross-li
C:Function:
A:Description: minor structural component of extracellular basement membrane
C:Superfamily: collagen alpha 1(IV) chain
C:Keywords: Airport syndrome; basement membrane; coiled coil; extracellular matrix; glyco
F,1-26/Domain: signal sequence #status predicted <SIG>
F,27-1691/Product: collagen alpha 5(IV) chain, renal splice form #status predicted <MATI>
F,27-1264,1271-1691/Product: collagen alpha 5(IV) chain, leukocyte splice form #status F
F,42-462/Domain: amino-terminal nonhelical, NCL #status predicted <NC2>
F,173-1573/Domain: collagen IV carboxyl-terminal repeat <CT1>
F,1583-1687/Domain: collagen IV carboxyl-terminal repeat <CT2>
F,125/Binding site: carbohydrate (Asn) (covalent) #status predicted
F,1482-1570,1515-1573/Disulfide bonds: (#status predicted
F,1533-1638-1687/Disulfide bonds: #status predicted
F,1592-1684,1626-1687/Disulfide bonds: (#or 1592-1687, 1626-1684) #status predicted
Query Match 75.0% Score 59; DB 1; Length 1691;
Best Local Similarity 73.3%; Pred. No. 0.04;
Matches 11; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
Qy 1 GYGGDKGNPGWGCAP 15
Db 1285 GIKGEKGNPGQGPCLP 1299

RESULT 5

A55267
collagen alpha 5(IV) chain - dog (fragment)
C:Species: Canis lupus familiaris (dog)
C:Date: 06-Feb-1995 #sequence_revision 06-Feb-1995 #text_change 13-Aug-1999
C:Accession: A55267
R,Zheng, K.; Thorner, P.S.; Marrano, P.; Baumal, R.; McInnes, R.R.
Proc. Natl. Acad. Sci. U.S.A. 91, 3969-3973, 1994
A>Title: Canine X chromosome-linked hereditary nephritis: a genetic model for human X-11
A:Reference number: A55267; MUID:94224868
A:Accession: A55267
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-754 <ZHE>
A,Cross-references: GB:U07888; NID:9469547; PIDN:AAB60258.1; PID:9469548
C:Superfamily: collagen alpha 1(IV) chain

Query Match 71.7% Score 66; DB 2; Length 754;
Best local Similarity 66.7%; Pred. No. 0.05; Matches 10; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GYGGDKGNPGWGCAP 15
Db 355 GIKGERENPGQGPCLP 369

RESULT 6

A33526
collagen alpha 2(IV) chain precursor - mouse
C:Species: Mus musculus (house mouse)
C:Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 13-Aug-1999
C:Accession: A33526; A24432; D24432; A25066; B23564; S19081; B29301; A24628; B32
R,Saus, J.; Quinones, S.; Mackrell, A.; Blumberg, B.; Muthukumaran, G.; Pihlajaniemi, T.
J. Biol. Chem. 264, 6318-6324, 1989
A>Title: The complete primary structure of mouse alpha-2(IV) collagen. Alignment with mo
A:Reference number: A33526; MUID:89197933
A:Accession: A33526

A,Molecule type: mRNA
A,Residues: 1-1707 <SAU>
A,Cross-references: EMBL:J04695; NID:9556298; PIDN:AAA50293.1; PID:9556299
R,Kurkinen, M.; Bernard, M.P.; Barlow, D.P.; Chow, L.T.
Nature 317, 177-179, 1985
A>Title: Characterization of 54-, 123- and 182-base-pair exons in the mouse alpha-2(I
A,Reference number: A39367; MUID:85296379
A:Accession: A24432
A,Molecule type: mRNA
A,Residues: 967-1096, 'G', 1098-1109 <KUL>
A,Cross-references: EMBL:X02896; NID:950263; PIDN:CAA26555.1; PID:9899326
A:Note: the authors translated the codon AAC for residue 964 as Lys
A:Accession: D24432
A,Molecule type: DNA
A,Residues: 964-1096 'G', 1098-1109 <KU2>
A,Cross-references: EMBL:X02899
R,Schwarz, U.; Schuppan, D.; Oberbaeumer, I.; Glanville, R.W.; Deutzmann, R.; Impl,
Eur. J. Biochem. 157, 49-56, 1985
A>Title: Structure of mouse type IV collagen. Amino-acid sequence of the C-terminal 5
A,Reference number: A25066; MUID:86220192
A,Molecule type: mRNA
A,Residues: 970-1480 <SC1>
A,Cross-references: EMBL:X04647
A:Accession: B25066
A,Molecule type: protein
A,Residues: 979-1058-1065-1101-1105-1222-1226-1310-1326-1335-1351-1480 <SC2>
R,Vogeli, G.; Horn, E.; Carter, J.; Raytes, P.S.
FEBS Lett. 206, 29-32, 1986
A:Title: Proposed alignment of helical interruptions in the two subunits of the basem
A,Reference number: A24364; MUID:87005245
A:Accession: A24364
A,Molecule type: mRNA
A,Residues: 1041-1050, 'R', 1052-1170, 'S', 1172-1178, 'R', 1180-1240, 'E', 1242-1327, 'A', 132
A,Cross-references: EMBL:X04410; NID:950240; PIDN:CA27998.1; PID:929678
A,Residues: 1466-1622, 'H', 1624-1707 <KA1>
R,Kurkinen, M.; Condron, M.R.; Blumberg, B.; Barlow, D.P.; Quinones, S.; Saus, J.; Pih
FEMS Lett. 262, 8496-8499, 1987
A>Title: Extensive homology between the carboxyl-terminal peptides of mouse alpha-1(I
A:Reference number: A94680; MUID:87250460
A:Accession: B29301
A,Molecule type: mRNA
A,Residues: 1481-1707 <KUR>
A,Cross-references: EMBL:MI5833; NID:9192284; PIDN:AAA37341.1; PID:9387116
R,Schwarz-Maggolen, U.; Oberbaeumer, I.; Kuehn, K.
FEBS Lett. 208, 203-207, 1986
A>Title: cDNA and protein sequence of the NCL domain of the alpha-2-chain of collagen
A:Reference number: A24628; MUID:87054581
A:Accession: A24628
A,Molecule type: DNA
A,Residues: 1480-1572, 'L', 1574-1622, 'H', 1624-1707 <SCH>
A,Cross-references: EMBL:X04647
R,Kaytes, P.; Wood, L.; Theriault, N.; Kurkinen, M.; Vogeli, G.
J. Biol. Chem. 263, 19274-19277, 1988
A>Title: Head-to-head arrangement of murine type IV collagen genes.
A,Reference number: A92702; MUID:89066738
A:Accession: B32003
A,Molecule type: DNA
A,Residues: 1-33 <KA2>
A,Cross-references: EMBL:J04448; NID:9192666; PIDN:AAA37438.1; PID:9192667
R,Burbelo, P.D.; Martin, G.R.; Yamada, Y.
Proc. Natl. Acad. Sci. U.S.A. 85, 6679-6682, 1988
A>Title: Alpha1(IV) and alpha2(IV) collagen genes are regulated by a bidirectional pr
A:Reference number: A94220; MUID:89071759
A:Accession: B31766
A,Molecule type: DNA

A; Residues: 1-60 <SUR>
A; Cross-references: EMBL:M23333
R; Weber, S.; Engel, J.; Wiedemann, H.; Glanville, R.W.; Timpl, R.
Eur. J. Biochem. 139, 401-410 1984
A; Title: Subunit structure and assembly of the globular domain of basement-membrane collagen
A; Reference number: S1701; MUID:88132058
A; Accession: S19086
A; Molecule type: protein
A; Residues: 1475-1481, X, 1483-1487 <WEB>
C; Genetics:
A; Intros: 15/2; 33/3; 963/1; 1003/3; 1064/3; 1085/3
C; Superfamily: collagen alpha 1(IV) chain
C; Keywords: basement membrane; cell binding; coiled coil; disulfide bond; duplication; e
F; 1-28/Domain: signal sequence #status predicted <SIG>
C; Domains: 7S #status predicted <7SD>
F; 29-171/Domain: collagenous #status predicted <COL>
F; 172-580/Domain: cell attachment (R-G-D) motif
F; 173-362/Region: cell attachment (R-G-D) motif
F; 174-781/Region: cell attachment (R-G-D) motif
F; 884-886/Region: cell attachment (R-G-D) motif
F; 955-967/Region: cell attachment (R-G-D) motif
F; 123-125/Region: cell attachment (R-G-D) motif
F; 147-149/Region: cell attachment (R-G-D) motif
F; 181-181/Domain: carboxyl-terminal nonhelical, NCL #status predicted <NCI1>
F; 1590-1707/Domain: repeat NCL #status predicted <NCI1>
F; 42-47,51,53-481,483/Disulfide bonds: interchain #status predicted
F; 188-192/Binding site: carbohydrate (Asn) (covalent) #status predicted
F; 656-676-154-1550,1653-1660/Disulfide bonds: #status predicted
F; 985, 988, 997, 1003, 1028, 1031, 1067, 1101, 1113, 1119, 1143, 1170, 1200, 1242, 1305, 1368, 1391/Modifications
F; 985, 988, 997, 1003, 1028, 1031, 1067, 1101, 1113, 1119, 1143, 1170, 1200, 1242, 1305, 1368, 1391/Modifications
F; 1109, 1012, 1018, 1021, 1024, 1037, 1040, 1043, 1046, 1052, 1058, 1070, 1098, 1110, 1128, 1140, 1149, 1177, 1383, 1386, 1401, 1408, 1420, 1423, 1429, 1444, 1465, 1468, 1471, 1477/Modified site: hydroxyproline
Query Match 70.7%; Score 65; DB 2; Length 170;
Best Local Similarity 66.7%; Pred. No. 0.16;
Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
Oy 1 GVKGDKGNPGWPGP 15
Db 1165 GIPSDKGDGPGP 1179

RESULT 7

CGH2B
collagen alpha 2(IV) chain precursor - human
N; Alternate names: procollagen alpha 2(IV) chain
C; Species: Homo sapiens (man)
C; Date: 07-Jun-1990 #sequence revision 03-Oct-1995 #text change 22-Jun-1999
C; Accession: A3204; S0007; S02624; S00246; S17678; S16911; B32117; S16877; S00165; S39
R; Hostikka, S.L.; Tryggvason, K.
J. Biol. Chem. 263, 19418-19493, 1988
A; Title: The complete primary structure of the alpha2 chain of human type IV collagen
A; Reference number: A32024; MUID:89066769
A; Accession: S12024
A; Molecule type: mRNA
A; Residues: 1-1712 <HOS1>
A; Cross-references: EMBL:J04210; EMBL:X05610; GB:M20753; NID:g29550; PIDN:CAA29098.1; PII: FEB5 Lett. 215, 281-286, 1987
R; Hostikka, S.L.; Kurkinen, M.; Tryggvason, K.
A; Title: Nucleotide sequence coding for the human type IV collagen alpha-2 chain cDNA related region. Note: 1399-1le was also found
A; Reference number: S00007; MUID:87219158
A; Molecule type: mRNA
A; Residues: 1554-1598, V, 1400-1712 <HOS2>
A; Cross-references: EMBL:J04210; EMBL:X05610; GB:M20753; NID:g29550; PIDN:CAA29098.1; PII: R; MacKnight, R.S.; Benson, V.A.; Lovello, K.T.; van der Rest, M.; Fietzek, P.P.
A; Title: Isolation and characterization of pepsin-solubilized human basement membrane
A; Reference number: S39615; MUID:94038863
A; Accession: S39615
A; Molecule type: protein
A; Residues: 407-570 <EBL>
R; MacKnight, R.S.; Benson, V.A.; Lovello, K.T.; van der Rest, M.; Fietzek, P.P.
A; Title: Isolation and characterization of pepsin-solubilized human basement membrane
A; Reference number: S16910; MUID:84053346
A; Accession: S16912
A; Molecule type: protein
A; Residues: 490-492, X, 494-496, 675-677, 'G', 679-680, 'G', 682, 684-685, 'P' <MAC>
A; Experimental source: Placenta
R; Glanville, R.W.; Rauter, A.
Hoppe-Seyler's Z. Physiol. Chem. 362, 943-951, 1981

A; Accession: S02624
A; Status: not compared with conceptual translation
A; Molecule type: DNA
A; Residues: 1347-1350; 1377-1383; 1426-1432; 1665-1711; 1529-1535; 1625-1630 <HOS3>
A; Note: complete nucleotide sequence not shown
R; Brazil, D.; Pollner, R.; Oberbaumer, I.; Kuehn, K.
Eur. J. Biochem. 172, 35-42, 1988
A; Title: Human basement membrane collagen (type IV): the amino acid sequence of the a
A; Reference number: S00246; MUID:88151998
A; Accession: S00246
A; Molecule type: mRNA
A; Residues: 1-682, 'G', 684-1043

R; Oberbaumer, I.
A; Reference number: S17678
A; Accession: S17678
A; Molecule type: mRNA
A; Residues: 1-470, 'P', 472-682, 'G', 684-1043 <OBE>
R; Cross-references: EMBL:X0562; NID:930075; PIDN:CAA29076.1; PID:930076
R; Poeschl, E.; Pollner, R.; Kuehn, K.
EMBO J. 7, 2687-2695, 1988
A; Title: The genes for the alpha1(IV) and alpha2(IV) chains of human basement membran
A; Reference number: S02738; MUID:89030632
A; Accession: S16911
A; Status: translation not shown
A; Molecule type: DNA
A; Residues: 1-33 <POE>
A; Cross-references: EMBL:X12784; GB:M36963; NID:930072; PIDN:CAA31275.1; PID:930073
R; Soininen, R.; Huotari, M.; Hostikka, S.L.; Prockop, D.J.; Tryggvason, K.
J. Biol. Chem. 263, 17217-17220, 1988
A; Title: The structural genes for alpha1 and alpha2 chains of human type IV collagen
A; Reference number: A92690; MUID:89034231
A; Accession: B32117
A; Molecule type: DNA
A; Residues: 1-33 <S012>
A; Cross-references: EMBL:J04217; EMBL:J05039; NID:9180759; PIDN:AAA3097.1; PID:95532
R; Soininen, R.; Huotari, M.; Ganguly, A.; Prockop, D.J.; Tryggvason, K.
J. Biol. Chem. 264, 13565-13571, 1989
A; Title: Structural organization of the gene for the alpha-1 chain of human type IV collagen
A; Reference number: S16876; MUID:89340433
A; Accession: S16877
A; Molecule type: nucleic acid
A; Residues: nucleic acid sequence not shown; translation not shown
A; Molecule type: DNA
A; Residues: 1-33 <S012>
A; Cross-references: EMBL:J04217; NID:g180759; PIDN:AAA53097.1; PID:9553233; EMBL:J050
R; Siebold, B.; Qian, R.Q.; Glanville, R.W.; Hofmann, H.; Deutzmann, R.; Kuehn, K.
Bur. J. Biochem. 108, 559-575, 1987
A; Title: Construction of a model for the aggregation and cross-linking region (7S dom
A; Region: is region, is region
A; Reference number: S00165; MUID:88029476
A; Accession: S00165
A; Molecule type: protein
A; Residues: 37-247 <S1E1>
A; Note: the sequence from Fig. 4 is inconsistent with that from Fig. 3 in having 175-
R; Ebile, J.A.; Golbik, R.; Mann, K.; Kuehn, K.
EMBO J. 12, 4795-4802, 1993
A; Title: Cross-1-beta-1 integrin recognition site of the basement membrane collagen
A; Reference number: S39614; MUID:94038863
A; Accession: S39615
A; Molecule type: protein
A; Residues: 407-570 <EBL>
R; MacKnight, R.S.; Benson, V.A.; Lovello, K.T.; van der Rest, M.; Fietzek, P.P.
A; Title: Isolation and characterization of pepsin-solubilized human basement membrane
A; Reference number: S16910; MUID:84053346
A; Accession: S16912
A; Molecule type: protein
A; Residues: 490-492, X, 494-496, 1983
A; Title: Isolation and characterization of pepsin-solubilized human basement membrane
A; Reference number: S16910; MUID:84053346
A; Accession: S16912
A; Molecule type: protein
A; Residues: 490-492, X, 494-496, 675-677, 'G', 679-680, 'G', 682, 684-685, 'P' <MAC>
A; Experimental source: Placenta
R; Glanville, R.W.; Rauter, A.
Hoppe-Seyler's Z. Physiol. Chem. 362, 943-951, 1981

Query Match		Score	DB	Length	Best Local Similarity	Pred.	No.	DB 1;	Length 1712;
Accession:					73.38;		0.23;		
A;Title: Pepsin fragments of human placental basement-membrane collagens showing interrupted sequence number: S16908; MUID:82005835	A;Reference number: S01450; MUID:88085168								
A;Molecule type: protein	A;Molecule type: mRNA								
A;Accession: S01450	A;Accession: S02550								
A;Residues: 1040, 'L', 1042-1398, 'V', 1400-1418, 'M', 1420-1635, 'V', 1637-1712 <KIL>	A;Residues: 1480-1535,1545-1614,1617-1662, 'H', 1664-1700, 'G', 1705-1708,1710-1712 <STE2>								
A;Cross-references: EMBL:M24766; NID:5537328; PIDN:AAA52043;1; PID:9537329	R;Siebold, B.; Deutzeckmann, R.; Kuhn, K.								
R;Myers, J.C.; Howard, P.S.; Jeelen, A.M.; Dion, A.S.; Macarak, E.J.	Eur. J. Biochem. 176, 617-624, 1988								
A;Title: The arrangement of intra- and intermolecular disulfide bonds in the carboxyterminal region of the alpha-2(IV) collagen chain and chromosomal localization	A;Title: The arrangement of intra- and intermolecular disulfide bonds in the carboxyterminal region of the alpha-2(IV) collagen chain and chromosomal localization								
A;Reference number: S02550; MUID:8900112	A;Reference number: S02550; MUID:8900112								
A;Accession: S02550	A;Accession: S02550								
A;Molecule type: protein	A;Molecule type: protein								
A;Residues: 1480-1535,1545-1614,1617-1662, 'H', 1664-1700, 'G', 1705-1708,1710-1712 <STE2>	C;Species: Gallus gallus (chicken)								
A;Note: the sequence form Fig. 7 is inconsistent with that shown in Fig. 11 in having 17 residues at position 1480-1535,1545-1614,1617-1662, 'H', 1664-1700, 'G', 1705-1708,1710-1712 <STE2>	C;Date: 22-Nov-1993 #sequence-revision 09-Mar-1996 #text_change 20-Sep-1999								
R;Myers, J.C.; Howard, P.S.; Jeelen, A.M.; Dion, A.S.; Macarak, E.J.	C;Accession: S23296; C3493; A29032; B28754; A18856; B18886; A28518; R;Ninomiya, Y.; Castagnola, P.; Gerecke, D.; Gordon, M.K.; Jacenko, O.; Luvalle, P.; maguchi, N.; Olsen, B.R.								
J. Biol. Chem. 262, 9221-9228, 1987	In Extracellular Matrix Genes, Sandell L.J. and Boyd C.D., eds., PP.79-114 Academic								
A;Title: Duplication of type IV collagen COOH-terminal repeats and species-specific expression	A;Title: The molecular biology of collagens with short triple-helical domains.								
A;Reference number: A27114; MUID:87250571	A;Reference number: S22243								
A;Molecule type: mRNA	A;Molecule type: DNA								
A;Residues: 1486-1574, 'I', 1576-1712 <WYE>	A;Residues: 1-677 <NI1>								
A;Cross-references: EMBL:J02260; NID:1180425; PIDN:AAA58422;1; PID:9180426	R;Nishimura, I.; Muragaki, Y.; Olsen, B.R.								
C;Comment: Prolines and lysines at the third position of the tripeptide repeating unit are ed and subsequently O-glycosylated.	J. Biol. Chem. 264, 2003-2004, 1989								
C;Genetics:	A;Status: not compared with conceptual translation								
A;Gene: GDB:COL4A2	A;Accession: S22296								
A;Cross-references: GDB:119792; OMIM:120090	A;Molecule type: mRNA								
A;Position: 13q34-13q34	A;Accession: C349493								
A;Introns: 15/2; 33/3; 1347/1; 1380/1; 1429/1; 1468/1; 1532/1; 1527/3 #status incomplete	A;Residues: 1-174 <NTS>								
A;Note: the alpha 1(IV) and alpha 2(IV) chain genes are encoded on opposite strands with different orientations.	A;Cross-references: EMBL:M28660; NID:9211625; PIDN:AAA48709;1; PID:9211626								
A;Complex: Type IV collagen is a heterotrimer of two alpha 1(IV) chains (see PIR:CCHB4B) domains (with disulfide and desmosine cross-links), dimeric associations among trimer can be formed by the C-terminal nonhelical domain.	R;McCormick, D.; van der Rest, M.; Goodship, J.; Lozano, G.; Ninomiya, Y.; Olsen, B.R.								
A;Cross-references: GDB:119792; OMIM:120090	Proc. Natl. Acad. Sci. U.S.A. 84, 4044-4048, 1987								
A;Position: 13q34-13q34	A;Title: Structure of the glycosaminoglycan domain in the type IX collagen-proteoglycan complex.								
A;Introns: 15/2; 33/3; 1347/1; 1380/1; 1429/1; 1468/1; 1532/1; 1527/3 #status incomplete	A;Accession: A29032								
A;Note: the alpha 1(IV) and alpha 2(IV) chain genes are encoded on opposite strands with different orientations.	A;Molecule type: mRNA								
A;Complex: Type IV collagen is a heterotrimer of two alpha 1(IV) chains (see PIR:CCHB4B) domains (with disulfide and desmosine cross-links), dimeric associations among trimer can be formed by the C-terminal nonhelical domain.	A;Accession: C29032; MUID:87231947								
A;Cross-references: GDB:119792; OMIM:120090	A;Title: Structure of the glycosaminoglycan domain in the type IX collagen-proteoglycan complex.								
A;Position: 13q34-13q34	A;Accession: A29032								
A;Introns: 15/2; 33/3; 1347/1; 1380/1; 1429/1; 1468/1; 1532/1; 1527/3 #status incomplete	A;Residues: 147-165,170-180, 'X', 182-188 <MC2>								
A;Note: the alpha 1(IV) and alpha 2(IV) chain genes are encoded on opposite strands with different orientations.	R;Lozano, G.; Olsen, B.R.								
A;Complex: Type IV collagen is a heterotrimer of two alpha 1(IV) chains (see PIR:CCHB4B) domains (with disulfide and desmosine cross-links), dimeric associations among trimer can be formed by the C-terminal nonhelical domain.	Submitted to the EMBL Data Library, April 1990								
A;Cross-references: GDB:119792; OMIM:120090	A;Reference number: S22062								
A;Position: 13q34-13q34	A;Accession: B28754								
A;Introns: 15/2; 33/3; 1347/1; 1380/1; 1429/1; 1468/1; 1532/1; 1527/3 #status incomplete	A;Molecule type: protein								
A;Note: the alpha 1(IV) and alpha 2(IV) chain genes are encoded on opposite strands with different orientations.	A;Accession: A29032								
A;Complex: Type IV collagen is a heterotrimer of two alpha 1(IV) chains (see PIR:CCHB4B) domains (with disulfide and desmosine cross-links), dimeric associations among trimer can be formed by the C-terminal nonhelical domain.	A;Residues: 147-165,170-180, 'X', 182-188 <MC2>								
A;Cross-references: GDB:119792; OMIM:120090	R;Lozano, G.; Olsen, B.R.								
A;Position: 13q34-13q34	Proc. Natl. Acad. Sci. U.S.A. 82, 4050-4054, 1985								
A;Introns: 15/2; 33/3; 1347/1; 1380/1; 1429/1; 1468/1; 1532/1; 1527/3 #status incomplete	A;Title: A distinct class of vertebrate collagen genes encodes chicken type IX collagen.								
A;Note: the alpha 1(IV) and alpha 2(IV) chain genes are encoded on opposite strands with different orientations.	A;Reference number: A94705; MUID:85216631								
A;Complex: Type IV collagen is a heterotrimer of two alpha 1(IV) chains (see PIR:CCHB4B) domains (with disulfide and desmosine cross-links), dimeric associations among trimer can be formed by the C-terminal nonhelical domain.	A;Accession: B28754								
A;Cross-references: GDB:119792; OMIM:120090	A;Molecule type: DNA								
A;Position: 13q34-13q34	A;Accession: B28754								
A;Introns: 15/2; 33/3; 1347/1; 1380/1; 1429/1; 1468/1; 1532/1; 1527/3 #status incomplete	A;Residues: 401, 'RA', 404-631, 'D', 633-677 <LOI>								
A;Note: the alpha 1(IV) and alpha 2(IV) chain genes are encoded on opposite strands with different orientations.	R;Lozano, G.; Olsen, B.R.								
A;Complex: Type IV collagen is a heterotrimer of two alpha 1(IV) chains (see PIR:CCHB4B) domains (with disulfide and desmosine cross-links), dimeric associations among trimer can be formed by the C-terminal nonhelical domain.	Submitted to the EMBL Data Library, April 1990								
A;Cross-references: GDB:119792; OMIM:120090	A;Reference number: S22062								
A;Position: 13q34-13q34	A;Accession: B28754								
A;Introns: 15/2; 33/3; 1347/1; 1380/1; 1429/1; 1468/1; 1532/1; 1527/3 #status incomplete	A;Molecule type: protein								
A;Note: the alpha 1(IV) and alpha 2(IV) chain genes are encoded on opposite strands with different orientations.	A;Accession: B28754								
A;Complex: Type IV collagen is a heterotrimer of two alpha 1(IV) chains (see PIR:CCHB4B) domains (with disulfide and desmosine cross-links), dimeric associations among trimer can be formed by the C-terminal nonhelical domain.	A;Residues: 401, 'RA', 404-631, 'D', 633-677 <LOI>								
A;Cross-references: GDB:119792; OMIM:120090	R;Lozano, G.; Olsen, B.R.								
A;Position: 13q34-13q34	Proc. Natl. Acad. Sci. U.S.A. 82, 4050-4054, 1985								
A;Introns: 15/2; 33/3; 1347/1; 1380/1; 1429/1; 1468/1; 1532/1; 1527/3 #status incomplete	A;Title: A distinct class of vertebrate collagen genes encodes chicken type IX collagen.								
A;Note: the alpha 1(IV) and alpha 2(IV) chain genes are encoded on opposite strands with different orientations.	A;Reference number: A94705; MUID:85216631								
A;Complex: Type IV collagen is a heterotrimer of two alpha 1(IV) chains (see PIR:CCHB4B) domains (with disulfide and desmosine cross-links), dimeric associations among trimer can be formed by the C-terminal nonhelical domain.	A;Accession: B28754								
A;Cross-references: GDB:119792; OMIM:120090	A;Molecule type: protein								
A;Position: 13q34-13q34	A;Accession: B28754								
A;Introns: 15/2; 33/3; 1347/1; 1380/1; 1429/1; 1468/1; 1532/1; 1527/3 #status incomplete	A;Residues: 401, 'RA', 404-631, 'D', 633-677 <LOI>								
A;Note: the alpha 1(IV) and alpha 2(IV) chain genes are encoded on opposite strands with different orientations.	R;Lozano, G.; Olsen, B.R.								
A;Complex: Type IV collagen is a heterotrimer of two alpha 1(IV) chains (see PIR:CCHB4B) domains (with disulfide and desmosine cross-links), dimeric associations among trimer can be formed by the C-terminal nonhelical domain.	Proc. Natl. Acad. Sci. U.S.A. 82, 4050-4054, 1985								
A;Cross-references: GDB:119792; OMIM:120090	A;Title: A distinct class of vertebrate collagen genes encodes chicken type IX collagen.								
A;Position: 13q34-13q34	A;Reference number: A94705; MUID:85216631								
A;Introns: 15/2; 33/3; 1347/1; 1380/1; 1429/1; 1468/1; 1532/1; 1527/3 #status incomplete	A;Accession: B28754								
A;Note: the alpha 1(IV) and alpha 2(IV) chain genes are encoded on opposite strands with different orientations.	A;Molecule type: protein								
A;Complex: Type IV collagen is a heterotrimer of two alpha 1(IV) chains (see PIR:CCHB4B) domains (with disulfide and desmosine cross-links), dimeric associations among trimer can be formed by the C-terminal nonhelical domain.	A;Accession: B28754								
A;Cross-references: GDB:119792; OMIM:120090	A;Residues: 401, 'RA', 404-631, 'D', 633-677 <LOI>								
A;Position: 13q34-13q34	R;Lozano, G.; Olsen, B.R.								
A;Introns: 15/2; 33/3; 1347/1; 1380/1; 1429/1; 1468/1; 1532/1; 1527/3 #status incomplete	Proc. Natl. Acad. Sci. U.S.A. 82, 4050-4054, 1985								
A;Note: the alpha 1(IV) and alpha 2(IV) chain genes are encoded on opposite strands with different orientations.	A;Title: A distinct class of vertebrate collagen genes encodes chicken type IX collagen.								
A;Complex: Type IV collagen is a heterotrimer of two alpha 1(IV) chains (see PIR:CCHB4B) domains (with disulfide and desmosine cross-links), dimeric associations among trimer can be formed by the C-terminal nonhelical domain.	A;Reference number: A94705; MUID:85216631								
A;Cross-references: GDB:119792; OMIM:120090	A;Accession: B28754								
A;Position: 13q34-13q34	A;Molecule type: protein								
A;Introns: 15/2; 33/3; 1347/1; 1380/1; 1429/1; 1468/1; 1532/1; 1527/3 #status incomplete	A;Accession: B28754								
A;Note: the alpha 1(IV) and alpha 2(IV) chain genes are encoded on opposite strands with different orientations.	A;Residues: 401, 'RA', 404-631, 'D', 633-677 <LOI>								
A;Complex: Type IV collagen is a heterotrimer of two alpha 1(IV) chains (see PIR:CCHB4B) domains (with disulfide and desmosine cross-links), dimeric associations among trimer can be formed by the C-terminal nonhelical domain.	R;Lozano, G.; Olsen, B.R.								
A;Cross-references: GDB:119792; OMIM:120090	Proc. Natl. Acad. Sci. U.S.A. 82, 4050-4054, 1985								
A;Position: 13q34-13q34	A;Title: A distinct class of vertebrate collagen genes encodes chicken type IX collagen.								
A;Introns: 15/2; 33/3; 1347/1; 1380/1; 1429/1; 1468/1; 1532/1; 1527/3 #status incomplete	A;Reference number: A94705; MUID:85216631								
A;Note: the alpha 1(IV) and alpha 2(IV) chain genes are encoded on opposite strands with different orientations.	A;Accession: B28754								
A;Complex: Type IV collagen is a heterotrimer of two alpha 1(IV) chains (see PIR:CCHB4B) domains (with disulfide and desmosine cross-links), dimeric associations among trimer can be formed by the C-terminal nonhelical domain.	A;Residues: 401, 'RA', 404-631, 'D', 633-677 <LOI>								
A;Cross-references: GDB:119792; OMIM:120090	R;Lozano, G.; Olsen, B.R.								
A;Position: 13q34-13q34	Proc. Natl. Acad. Sci. U.S.A. 82, 4050-4054, 1985								
A;Introns: 15/2; 33/3; 1347/1; 1380/1; 1429/1; 1468/1; 1532/1; 1527/3 #status incomplete	A;Title: A distinct class of vertebrate collagen genes encodes chicken type IX collagen.								
A;Note: the alpha 1(IV) and alpha 2(IV) chain genes are encoded on opposite strands with different orientations.	A;Reference number: A94705; MUID:85216631								
A;Complex: Type IV collagen is a heterotrimer of two alpha 1(IV) chains (see PIR:CCHB4B) domains (with disulfide and desmosine cross-links), dimeric associations among trimer can be formed by the C-terminal nonhelical domain.	A;Accession: B28754								
A;Cross-references: GDB:119792; OMIM:120090	A;Molecule type: protein								
A;Position: 13q34-13q34	A;Accession: B28754								
A;Introns: 15/2; 33/3; 1347/1; 1380/1; 1429/1; 1468/1; 1532/1; 1527/3 #status incomplete	A;Residues: 401, 'RA', 404-631, 'D', 633-677 <LOI>								
A;Note: the alpha 1(IV) and alpha 2(IV) chain genes are encoded on opposite strands with different orientations.	R;Lozano, G.; Olsen, B.R.								
A;Complex: Type IV collagen is a heterotrimer of two alpha 1(IV) chains (see PIR:CCHB4B) domains (with disulfide and desmosine cross-links), dimeric associations among trimer can be formed by the C-terminal nonhelical domain.	Proc. Natl. Acad. Sci. U.S.A. 82, 4050-4054, 1985								
A;Cross-references: GDB:119792; OMIM:120090	A;Title: A distinct class of vertebrate collagen genes encodes chicken type IX collagen.								
A;Position: 13q34-13q34	A;Reference number: A94705; MUID:85216631								
A;Introns: 15/2; 33/3; 1347/1; 1380/1; 1429/1; 1468/1; 1532/1; 1527/3 #status incomplete	A;Accession: B28754								
A;Note: the alpha 1(IV) and alpha 2(IV) chain genes are encoded on opposite strands with different orientations.	A;Residues: 401, 'RA', 404-631, 'D', 633-677 <LOI>								
A;Complex: Type IV collagen is a heterotrimer of two alpha 1(IV) chains (see PIR:CCHB4B) domains (with disulfide and desmosine cross-links), dimeric associations among trimer can be formed by the C-terminal nonhelical domain.	R;Lozano, G.; Olsen, B.R.								
A;Cross-references: GDB:119792; OMIM:120090	Proc. Natl. Acad. Sci. U.S.A. 82, 4050-4054, 1985								
A;Position: 13q34-13q34	A;Title: A distinct class of vertebrate collagen genes encodes chicken type IX collagen.								
A;Introns: 15/2; 33/3; 1347/1; 1380/1; 1429/1; 1468/1; 1532/1; 1527/3 #status incomplete	A;Reference number: A94705; MUID:85216631								
A;Note: the alpha 1(IV) and alpha 2(IV) chain genes are encoded on opposite strands with different orientations.	A;Accession: B28754								
A;Complex: Type IV collagen is a heterotrimer of two alpha 1(IV) chains (see PIR:CCHB4B) domains (with disulfide and desmosine cross-links), dimeric associations among trimer can be formed by the C-terminal nonhelical domain.	A;Residues: 401, 'RA', 404-631, 'D', 633-677 <LOI>								
A;Cross-references: GDB:119792; OMIM:120090	R;Lozano, G.; Olsen, B.R.								
A;Position: 13q34-13q34	Proc. Natl. Acad. Sci. U.S.A. 82, 4050-4054, 1985								
A;Introns: 15/2; 33/3; 1347/1; 1380/1; 1429/1; 1468/1; 1532/1; 1527/3 #status incomplete	A;Title: A distinct class of vertebrate collagen genes encodes chicken type IX collagen.								
A;Note: the alpha 1(IV) and alpha 2(IV) chain genes are encoded on opposite strands with different orientations.	A;Reference number: A94705; MUID:85216631								
A;Complex: Type IV collagen is a heterotrimer of two alpha 1(IV) chains (see PIR:CCHB4B) domains (with disulfide and desmosine cross-links), dimeric associations among trimer can be formed by the C-terminal nonhelical domain.	A;Accession: B28754								
A;Cross-references: GDB:119792; OMIM:120090	A;Molecule type: protein								
A;Position: 13q34-13q34	A;Accession: B28754								
A;Introns: 15/2; 33/3; 1347/1; 1380/1; 1429/1; 1468/1; 1532/1; 1527/3 #status incomplete	A;Residues: 401, 'RA', 404-631, 'D', 633-677 <LOI>								
A;Note: the alpha 1(IV) and alpha 2(IV) chain									

A;Accession: A28518
A;Molecule type: protein
A;Residues: 156-159, 'X', 161-166, 'X', 168-173, 'X', 175-178 <HUB>
A;Accession: S23813
A;Molecule type: protein
A;Residues: 170-180, 'X', 182-184 <VAN>
A;Note: evidence for aldimine cross-linkage of 190-Lys to collagen alpha 1(II) chain is
R;Maine, R.; van der Rest, M.; Niizumiya, Y.; Olsen, B.R.
Ann. N.Y. Acad. Sci. 460, 38-46, 1985
A;Title: The structure of type IX collagen.
A;Reference number: S22238; MUID:86185164
A;Molecule type: protein
A;Residues: 542-567 <MAY>
C;Introns: 427/3; 439/3; 454/3; 465/3; 514/3; 533/1; 596/1; 622/1
C;Superfamily: unassigned collagens
C;Keywords: chondroitin sulfate proteoglycan; coiled coil; extracellular matrix; glycoprotein
F;1-21/Domain: signal sequence #status predicted <SIG>
F;22-67/Product: collagen alpha 2(XI) chain #status predicted <MAT>
F;25-161/Domain: collagenous COL3 #status predicted <COL3>
F;62-178/Domain: non-collagenous NC3 #status predicted <NC3>
F;175-517/Domain: collagenous COL2 #status predicted <COL2>
F;548-547/Domain: non-collagenous NC2 #status predicted <COL2>
F;548-662/Domain: collagenous COL1 #status predicted <COL1>
F;663-677/Domain: non-collagenous NC1 #status predicted <NC1>
F;158-178/Modified site: hydroxyproline (Pro) #status experimental
F;167-Binding site: hydroxyproline (Ser) (covalent) #status experimental
F;181/Modified site: 5-hydroxylysine (Lys) #status experimental
F;181/Binding site: carbohydrate (Lys) (covalent) #status experimental
F;190/Modified site: allysine (Lys) #status predicted
Query Match Score 68.5%; Best Local Similarity 66.7%; Score 63; DB 2; Length 677;
Matches 10; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
Qy 1 GVKGKGNPCPGPAG 15
Db 46 GIDGDKGSPGAPGSP 60

RESULT 9

A54849

collagen alpha 1(VII) chain precursor - human
N;Alternate names: procollagen alpha 1(VII) chain
C;Species: Homo sapiens (man)
C;Date: 04-Nov-1994 #sequence_revision 04-Nov-1994 #text_change 20-Sep-1999
C;Accession: A54849; PH0844; Sl16316; 156328; A30296; 184686
R;Christiano, A.M.; Greenspan, D.S.; Lee, S.; Uitto, J.
J. Biol. Chem. 269, 20266-20262, 1994
A;Title: Cloning of human type VII collagen. Complete primary sequence of the alpha1(VII) chain
A;Reference number: A54849; MUID:94327588
A;Accession: A54849
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-294 <CHR>
A;Cross-references: GB:LU2870; PIDN:9987124; PIDN:AAA7438.1; PID:9987125
A;Residues: 340-475, 'RALSTASHSTLWRTRWPHCNRGSHTRACEPCPNPASHRAARAG', 524-528, 'C'
R;Tanaka, T.; Takahashi, K.; Furukawa, F.; Iimamura, S.
Biochem. Biophys. Res. Commun. 183, 958-963, 1992
A;Title: Molecular cloning and characterization of type VII collagen cDNA.
A;Reference number: PH0844; MUID:92231902
A;Accession: PH0844
A;Molecule type: mRNA
A;Residues: 'EPR', 340-475, 'RALSTASHSTLWRTRWPHCNRGSHTRACEPCPNPASHRAARAG', 524-528, 'C'
A;Cross-references: DDBJ:DI1152; DDBJ:DI3694; PIDN:BA02853.1; PID:9453699
A;Experimental source: keratinocyte
A;Note: the authors translated the codon ACC for residues 394 and 397 as Tyr
R;Parente, M.G.; Chung, L.C.; Ryymänen, J.; Woodley, D.T.; Wynn, K.C.; Bauer, E.A.; Mat
Proc. Natl. Acad. Sci. U.S.A. 88, 6931-6935, 1991

A;Title: Human type VII collagen: cDNA cloning and chromosomal mapping of the gene.
A;Reference number: S16316; MUID:9134380
A;Accession: S16316
A;Molecule type: mRNA
A;Residues: 815-892, 'E', 894-1439 <PAR>
A;Cross-references: GB:MO5158; GB:S49017; PID:9180914; PIDN:AAA96439.1; PID:9180915
A;Experimental source: keratinocyte
R;Gammon, W.R.; Abernethy, M.L.; Padilla, K.M.; Prisavanh, P.S.; Cook, M.E.; Wright, J. Invest. Dermatol. 99, 691-696, 1992
A;Title: Noncollagenous (NC1) domain of collagen VII resembles multidomain adhesion proteins
A;Reference number: I56328; MUID:93107742
A;Accession: I56328
A;Title: Cleavage of type VII collagen by interstitial collagenase and type IV collagen
A;Reference number: A30296; MUID:89139437
A;Accession: A30296
A;Molecule type: protein
A;Residues: 'A', 1240-1246, 'G', 1248-1250, 'XE', 1253-1255, 'Q', 1257, 'E', 2032, 'C', 2034-204
A;Note: two reported peptides cannot be reliably located
R;Greenspan, D.S.
Hum. Mol. Genet. 2, 273-278, 1993
A;Title: The carboxyl-terminal half of type VII collagen, including the non-collagenous
A;Reference number: I48103; MUID:93271985
A;Accession: I04686
A;Status: preliminary; translated from GB/EMBL/DDJB
A;Molecule type: mRNA
A;Residues: 'A', 1240-1246, 'G', 1248-1250, 'XE', 1253-1255, 'Q', 1257, 'E', 2032, 'C', 2034-204
A;Cross-references: GB:LU6862; PID:9388713; PIDN:AAA89196.1; PID:9388714
R;Christiano, A.M.; Ryynänen, M.; Uitto, J.
Proc. Natl. Acad. Sci. U.S.A. 91, 3549-3553, 1994
A;Title: Dominant dystrophic epidermolysis bullosa: identification of a GLY --> Ser substitution at position 1253 of the type VII collagen molecule
A;Reference number: A55255; MUID:94224777
A;Comments: annotation
C;Comment: Prolines and lysines at the third position of the tripeptide repeating unit
C;Genetics:
A;Gene: GDB:COL7A1; EBRI; EBDI; EB
A;Cross-references: GDB:128750; OMIM:120120
A;Map position: 3p21.3-3p21.3
A;Note: defects in this gene can result in dominant and recessive dystrophic epidermolysis bullosa
C;Complex: type VII collagen is probably a homotrimer
C;Function:
A;Description: structural component of extracellular polymer associated with anchoring
C;Superfamily: unassigned collagens; animal Kunitz-type proteinase inhibitor homology
C;Keywords: coiled coil; extracellular matrix; glycoprotein; hydroxylysine; hydroxyproline
F;1-16/Domain: signal sequence #status predicted <SIG>
F;17-1253/Domain: amino-terminal nonhelical #status predicted <NC1>
F;17-294/Product: collagen alpha 1(VII) chain #status predicted <MAT>
F;36-201/Domain: von Willebrand factor A repeat homology <WVAL>
F;231-318/Domain: fibronectin type III repeat homology <FN1>
F;327-413/Domain: fibronectin type III repeat homology <FN2>
F;414-502/Domain: fibronectin type III repeat homology <FN3>
F;509-593/Domain: fibronectin type III repeat homology <FN4>
F;598-683/Domain: fibronectin type III repeat homology <FN5>
F;686-771/Domain: fibronectin type III repeat homology <FN6>
F;1254-2703/Region: fibronectin type III repeat homology <FN7>
F;1334-1336/Region: cell attachment (R-G-D) motif
F;2008-2010/Region: cell attachment (R-G-D) motif
F;2553-2559/Region: cell attachment (R-G-D) motif
F;2784-2949/Domain: fibronectin type III repeat homology <FN8>
F;1052-1219/Domain: von Willebrand factor type A repeat homology <WVA2>
F;110-112/Region: cell attachment (R-G-D) motif
F;1189-1233/Region: cysteine/proline-rich
F;1254-2703/Region: interrupted heptapeptide
F;1334-1336/Region: cell attachment (R-G-D) motif
F;2008-2010/Region: cell attachment (R-G-D) motif
F;2553-2559/Region: cell attachment (R-G-D) motif
F;2784-2949/Domain: carboxyl-terminal nonhelical #status predicted <NC2>
F;2876-2929/Domain: animal Kunitz-type proteinase inhibitor homology <BPI>

F;337-786;1109/Binding site: carbohydrate (Asn) (covalent) #status predicted	C;Species: Homo sapiens (man)
F;2167-2176-2185-2188-2189-2190-2191/Modified site: 4-hydroxyproline (Pro) #status exper	C;Date: 28-Oct-1994 #sequence_revision 03-Oct-1995 #text_change 22-Jun-1999
F;2625-2631/Binding site: carbohydrate (Lys) (covalent) #status experimental	C;Accession: A54763; A43928; A44043; A45971; A3786
F;2634,2802,2804/Disulfide bonds: interchain #status predicted	R;Mariyama, M.; Leinonen, A.; Mochizuki, T.; Tryggvason, K.; Reeder, S.T.
Query Match, Best Local Similarity 80.0%; Score 63; DB 2; Length 294;	J. Biol. Chem., 269, 23013-23017, 1994
Matches 12; Conservative 0; Mismatches 3; Indels 0; Gaps 0;	A;Title: Complete primary structure of the human alpha3(IV) collagen chain. Coexpress
Qy 1 GVKGDKGNPGWPGAP 15	A;Reference number: A5763; MUID:9436994
Db 2386 GVKGDKGNPGWPGAP 2400	A;Accession: A54763
RESULT 10	A;Molecule type: mRNA
collagen - sea urchin (<i>Paracentrotus lividus</i>) (fragment)	A;Residues: 1-1670 <MAR>
C;Species: <i>Paracentrotus lividus</i> (common urchin)	A;Cross-references: GB:X80031; NID:9577563; PID:9577564
C;date: 17-Aug-1989 #sequence_revision 17-Aug-1989 #text_change 19-Jan-1996	A;Experimental source: kidney
R;Saitta, B.; Buttice, G.; Gambino, R.	R;Turner, N.; Mason, P.J.; Brown, R.; Fox, M.; Povey, S.; Rees, A.; Pusey, C.D.
Blochem. Biophys. Res. Commun., 158, 633-639, 1989	A;Title: Molecular cloning of the human Goodpasture antigen demonstrates it to be the
A;Title: Isolation of a putative collagen-like gene from the sea urchin <i>Paracentrotus lividus</i>	A;Reference number: A43928; MUID:92147878
A;reference number: A32249; MUID:89149773	A;Accession: A43928
A;Status: preliminary; not compared with conceptual translation	A;Residues: 1331-1524, 'I', 1526-1670 <UR>
A;Molecule type: DNA	A;Cross-references: GB:M81379
C;Status: preliminary	A;Experimental source: kidney
A;Residues: 1-290 <SAI>	R;Quinones, S.; Bernal, D.; Garcia-Sogo, M.; Elena, S.F.; Saus, J.
C;Superfamily: collagen alpha 2(I) chain; fibrillar collagen carboxyl-terminal homology	J. Biol. Chem., 267, 19780-19784, 1992
C;Keywords: coiled coil; extracellular matrix; glycoprotein; trimer; triple helix	A;Title: Exon/intron structure of the human alpha 3(IV) gene encompassing the Goodpasture antigen.
Query Match, Best Local Similarity 78.6%; Score 62; DB 2; Length 290; Matches 11; Conservative 1; Mismatches 2; Indels 0; Gaps 0;	A;Reference number: A44043; MUID:93015826
Qy 1 GVKGDKGNPGWPGAP 14	A;Accession: A44043
Db 76 GSKDQGNPGQPGA 89	A;Molecule type: mRNA
RESULT 11	A;Cross-references: GB:M92993; NID:9177895; PIDN:AAA21610.1; PID:9177896
S23809 collagen alpha 2(I) chain homolog - sea urchin (<i>Strongylocentrotus purpuratus</i>)	A;Residues: 1386-1670 <QUT>
C;Species: <i>Strongylocentrotus purpuratus</i> (purple urchin)	A;Sequence extracted from NCBI backbone (NCBIP:115597)
C;date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999	R;Quinones, S.; Bernal, D.; Garcia-Sogo, M.; Elena, S.F.; Saus, J.
C;Accession: S23809	J. Biol. Chem., 269, 12352-12358, 1994
R;Exposito, J.R.; d'Alessio, M.; Solursh, M.; Ramirez, F.	A;Reference number: A44738; MUID:9427734
J. Biol. Chem., 267, 15559-15562, 1992	A;Contents: annotation; erratum; correction to intronic sequence in A44043
A;Title: Sea urchin collagen evolutionarily homologous to vertebrate pro-alpha-2(I) coll	R;Bernal, D.; Quinones, S.; Saus, J.
A;Reference number: S23809; MUID:92348411	J. Biol. Chem., 268, 12090-12094, 1993
A;Status: preliminary	A;Title: The human mRNA encoding the Goodpasture antigen is alternatively spliced.
A;Molecule type: mRNA	A;Reference number: A45971; MUID:93280184
A;Residues: 1-144 <EXP>	A;Accession: A45971
A;Cross-references: EMBL:M92040; NID:9161335; PIDN:AAA30035.1; PID:9161336	A;Status: nucleic acid sequence not shown
C;Superfamily: collagen alpha 2(I) chain; fibrillar collagen carboxyl-terminal homology	A;Molecule type: mRNA
C;Keywords: coiled coil; extracellular matrix; glycoprotein; trimer; triple helix	A;Residues: 1427-1444 <BER>
F;1207-1414/domain: fibrillar collagen carboxyl-terminal homology <FCC>	A;Note: sequence extracted from NCBI backbone (NCBIP:133333); sequence incorrectly identified
Query Match, Best Local Similarity 78.6%; Score 62; DB 1; Length 144; Matches 11; Conservative 1; Mismatches 2; Indels 0; Gaps 0;	R;Morrison, K.E.; Mariyama, M.; Yang Peng, T.L.; Reeder, S.T.
Qy 1 GVKGDKGNPGWPGAP 14	A;Cross-references: GB:S55790; NID:9244118; PIDN:AAH19637.1; PID:9234419
Db 393 GSKDQGNPGQPGA 406	A;Comment: Prolines and lysines at the third position of the tripeptide repeating unit and subsequently O-glycosylated.
RESULT 12	A;Accession: A39786; MUID:91353570
CGH03B collagen alpha 3(IV) chain precursor, long splice form - human chain-long splice form	A;Accession: A39786
Alternative names: Goodpasture antigen; procollagen alpha 3(IV) chain-long splice form	A;Molecule type: mRNA

F;996-998/Region: cell attachment (R-G-D) motif
 F;1154-1156/Region: cell attachment (R-G-D) motif
 F;1306-1308/Region: cell attachment (R-G-D) motif
 C;Species: Homo sapiens (man)
 C;Accession: S23810; PQ0612; S08012
 R;Pan, T.C.; Zhang, R.Z.; Mattei, M.G.; Timpl, R.; Chu, M.L.
 Proc. Natl. Acad. Sci. U.S.A. 89, 6565-6569, 1992
 A;Title: Cloning and chromosomal location of human alpha1(XVI) collagen.
 A;Reference number: S23810; MID:92335339
 A;Accession: S23810
 A;Molecule type: mRNA
 A;Residues: 1-1603 <PAN>
 A;Cross-references: EMBL:M92542; NID:gi180757; PIDN:AAA58427.1; PID:gi180758
 A;Experimental source: skin fibroblasts
 R;Yamaguchi, N.; Kimura, S.; McBride, O.W.; Hori, H.; Yamada, Y.; Kanamori, T.; Yamakoshi, J. Biochem. 112, 856-863, 1992

Query Match 67.4%; Score 62; DB 1; Length 1670;
 Best Local Similarity 66.7%; Pred. No. 0.44;
 Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 1 GVKGDKGNPWPWPG 15
 Db 1189 GAKGDGRGAPGPGLP 1203

RESULT 13

B46345
 gene V protein - phage PRD1
 N;Alternate names: minor capsid protein V
 C;Species: Phage PRD1
 C;Accession: 17-Feb-1994 #sequence_revision 17-Feb-1994 #text_change 23-Jul-1999
 C;Accession: B46345; D46345
 R;Bamford, J.K.H.; Bamford, D.H.
 R;Virology 177, 445-51, 1990
 A;Title: Capsomeric proteins of bacteriophage PRD1, a bacterial virus with a membrane.
 A;Reference number: A46345; MID:90320115
 A;Molecule type: DNA
 A;Residues: 1-340 <BAM>
 A;Cross-references: GB:055568; NID:gi215745; PIDN:AAA32446.1; PID:gi215746
 A;Accession: B46345
 A;Molecule type: protein
 A;Residues: 2-6 <BAZ>
 C;Genetics:
 A;Gene: V
 C;Superfamily: phage PRD1 gene V protein
 C;Keywords: capsid protein

Query Match 66.3%; Score 61; DB 1; Length 340;
 Best Local Similarity 76.9%; Pred. No. 0.12; Matches 10; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GVKGDKGNPWPWPG 13
 Db 124 GIKGDKGDRGAPG 136

RESULT 14

S23810
 collagen alpha 1(XVI) chain precursor - human
 N;Alternate names: procollagen alpha 1(XVI) chain
 C;Species: Homo sapiens (man)
 C;Accession: S23810; PQ0612; S08012
 R;Leinonen, A.; Maruyama, M.; Mochizuki, T.; Tryggvason, K.; Reeders, S.T.
 J. Biol. Chem. 269, 26172-26177, 1994
 A;Title: Complete primary structure of the human type IV collagen alpha1(XVI) chain. C
 A;Reference number: A55360; MID:95014445
 A;Accession: A55360
 A;Status: nucleic acid sequence not shown
 A;Molecule type: mRNA
 A;Residues: 1-1690 <LEI>
 A;Cross-references: GB:X81053; NID:gi574805; PIDN:CAA56943.1; PID:gi574806
 R;Sugimoto, M.; Oohashi, T.; Yoshioka, H.; Matsuo, N.; Ninomiya, Y.

FBB3 Lett. 330, 122-128, 1993
A;Title: cDNA isolation and partial gene structure of the human alpha-4(IV) collagen chain
A;Reference number: S36854; MUID:9374047
A;Accession: S36854
A;Molecule type: DNA; mRNA
A;Cross-references: DDBJ:017391; NID:9440365; PIDN:AAA04214.1; PID:9457161
A;Experimental source: whole eye
R;Kamogata, Y.; Mattei, M.G.; Ninomiya, Y.
J. Biol. Chem. 267, 23753-23758, 1992
A;Title: Isolation and sequencing of cDNAs and genomic DNAs encoding the alpha4 chain of
A;Reference number: S28777; MUID:93054733
A;Accession: S28777
A;Molecule type: DNA
A;Residues: 1407-1424, 'G' 1426-1430, 'A' 1432-1439, 'L' 1441-1507 <KAM>
A;Cross-references: GB:LO1475; GB:LO1476
A;Note: the codons given for 1438-Asp (GAC) and 1443-Gly (GCA) are inconsistent with the
C;Comment: Prolines and Lysines at the third position of the tripeptide repeating unit
ed and subsequently O-glycosylated.
C;Genetics:
A;Gene: GDB:COLA4
A;Cross-references: GDB:132673; OMIM:120131
A;MM position: 2q35; 2q37
A;Introns: 39/1; 1406/1; 1445/1; 1508/1; 1603/3 #status incomplete
C;Complex: this minor type IV collagen is thought to form a heterotrimer of two alpha 3(I)
mon trimer amino-terminal domains (with disulfide and desmosine cross-links), dimeric
ter associations in the interrupted helical domain (with disulfide and desmosine cross-link)
C;Function:
A;Description: minor structural component of extracellular basement membrane in kidney gl
C;Superfamily: collagen alpha 1(IV) chain
C;Keywords: basement membrane; coiled coil; extracellular matrix; glycoprotein; hydroxyl
F:1-38/Domain: signal sequence #status predicted <SIG>
F:39-1690/Product: collagen alpha 4(IV) chain #status predicted <MAT>
F:39-61/Domain: amino-terminal nonhelical, NHI <NHI>
F:62-466/Region: interrupted helical
F:94-96/Region: cell attachment (R-G-D) motif
F:145-147/Region: cell attachment (R-G-D) motif
F:189-191/Region: cell attachment (R-G-D) motif
F:310-312/Region: cell attachment (R-G-D) motif
F:724-726/Region: cell attachment (R-G-D) motif
F:785-787/Region: cell attachment (R-G-D) motif
F:989-991/Region: cell attachment (R-G-D) motif
F:1212-121/Region: cell attachment (R-G-D) motif
F:1467-1690/Domain: carboxyl-terminal nonhelical, NCI <NCI>
F:1471-1569/Domain: collagen IV carboxyl-terminal repeat <CTR>
F:1579-1585/Domain: collagen IV carboxyl-terminal repeat <CTR>
F:47,52,55,57,266,400,460,492,494,668,790,828,1095,1131,1294,1317,1375,1407/disulfide bc
F:142,669/binding site: carbohydrate (Asn) (covalent) #status predicted <CH>
F:1480-1569,1513-1569/disulfide bonds: (or 1480,1569, 1513-1566) #status predicted
F:1522-1531,1632-1641/disulfide bonds: #status predicted
F:1588-1683,1622-1685/disulfide bonds: (or 1588-1686, 1622-1683) #status predicted
Query Match 66.3%; Score 61; DB 1; Length 1690;
Best Local Similarity 60.0%; Pred. No. 0.62; 4; Mismatches 2; Indels 0; Gaps 0;
Matches 9; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GVKGDKGNPGPAGA 15
Db 184 GIQGDRODPGLGGLP 198
RESULT 16
S16366 collagen alpha 2(IV) chain precursor - pig roundworm
C;Species: Ascaris suum (pig roundworm)
C;Accession: S16366 #sequence_revision 04-Dec-1992 #text_change 13-Aug-1999
C;Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 13-Aug-1999
C;Accession: S16366
R;Pettitt, J.; Kingston, I.B.
J. Biol. Chem. 266, 16149-16156, 1991
A;Title: The complete primary structure of a nematode alpha-2(IV) collagen and the parti
A;Reference number: S16366; MUID:9140768
A;Accession: S16366
A;Molecule type: mRNA
A;Residues: 1-1763 <BBI>
A;Cross-references: GB:MG7507; NID:9159648; PIDN:AAA18014.1; PID:9159649
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-623 <KAM>
A;Experimental source: basement membrane
A;Note: sequence extracted from NCBI backbone (NCBIP:118549)
C;Superfamily: collagen alpha 1(IV) chain
A;Cross-references: EMBL:U07974; NID:9520456; PIDN:AAA03409.1; PID:9537433
A;Genetics:
C;Species: Gallus gallus (chicken)
C;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 13-Aug-1999
C;Accession: I50696
R;Nah, H.D.; Niu, Z.; Adams, S.L.
J. Biol. Chem. 269, 16443-16448, 1994
A;Title: An alternative transcript of the chick type III collagen gene that does not
A;Reference number: A54041; MUID:94266842
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-310 <NHA>
A;Cross-references: EMBL:U07974; NID:9520456; PIDN:AAA03409.1; PID:9537433
A;Genetics:
C;Species: Coli3A1
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homolo
Query Match 65.2%; Score 60; DB 2; Length 310;
Best Local Similarity 71.4%; Pred. No. 0.16; DB 2; Length 310;
Matches 10; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
Qy 1 GVKGDKGNPGPAGA 14
Db 281 GYKGERGSPGPAGA 294
RESULT 18
A45137 collagen alpha 4(IV) chain - rabbit
C;Species: Oryctolagus cuniculus (domestic rabbit)
C;Accession: A45137
R;Kamogata, Y.; Mattei, M.G.; Ninomiya, Y.
J. Biol. Chem. 267, 23753-23758, 1992
A;Title: Isolation and sequencing of cDNAs and genomic DNAs encoding the alpha4 chain
A;Reference number: S28777; MUID:93054733
A;Accession: A45137
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-623 <KAM>
A;Title: Isolation and sequencing of cDNAs and genomic DNAs encoding the alpha4 chain
A;Experimental source: basement membrane
A;Note: sequence extracted from NCBI backbone (NCBIP:118549)

Query Match 65.2%; Score 60; DB 2; Length 623;
 Best Local Similarity 76.9%; Pred. No. 0.32; Matches 10; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Matches 10; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GVKGDKGNPGWPG 13
 Db 139 GIKDGDKGEGCSPG 151

RESULT 19
 S41067

collagen alpha 1(III) chain - rat
 C;Species: Rattus norvegicus (Norway rat)
 C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 13-Aug-1999
 C;Accession: S41067; A29905; S31924
 R;Glumoff, V.; Maekela, J.K.; Vuorio, E.
 Biochim. Biophys. Acta 1217, 41-48, 1994
 A;Title: Cloning of cDNA for rat pro alpha-1(III) collagen mRNA. Different expression patterns
 A;Reference number: S41067; MUID:94114571
 A;Accession: S41067
 A;Status: preliminary
 A;Molecule type: mRNA
 A;Residues: 1-636 <GLU>
 A;Cross-references: EMBL:X70369; NID:957915; PIDN:CAA49832.1; PID:957916
 R;Frankel, F.R.; Hsu, C.Y.J.; Myers, J.C.; Lin, E.; Lyttle, C.R.; Komm, B.; Mohn, K.
 DNA, 7, 347-354, 1988
 A;Title: Regulation of alpha-2(I), alpha-1 (III), and alpha-2(V) collagen mRNAs by estrogen
 A;Reference number: A29905; MUID:88296083
 A;Accession: A29905
 A;Molecule type: mRNA
 A;Residues: 308-482 <FRA>
 A;Cross-references: GB:A21354; NID:g203500; PIDN:AAA40942.1; PID:g203501
 R;Gumoff, V.; Maekela, J.K.; Vuorio, E.
 submitted to the EMBL data library, February 1993
 A;Reference number: S31924
 A;Accession: S31924
 A;Status: preliminary
 A;Molecule type: mRNA
 A;Residues: 2-636 <GLI2>
 A;Cross-references: EMBL:X70369
 C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology:
 C;Keywords: coiled coil; extracellular matrix; glycoprotein; trimer; triple helix
 F;44565/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match 65.2%; Score 60; DB 2; Length 636;
 Best Local Similarity 71.4%; Pred. No. 0.33; Matches 10; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Matches 10; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GVKGDKGNPGWPG 14
 Db 30 GVKGERGSPGGPA 43

RESULT 20
 I50594

collagen alpha 1(III) chain - chicken (fragment)
 C;Species: Gallus gallus (chicken)
 C;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 13-Aug-1999
 C;Accession: I50594
 R;Nish, H.D.; Mu, Z.; Adams, S.L. 1994
 J. Biol. Chem. 269, 16443-16448, 1994
 A;Title: An alternative transcript of the chick type III collagen gene that does not encode
 A;Reference number: A54041; MUID:94266842
 A;Accession: I50594
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: mRNA
 A;Residues: 1-1886 <NAH>
 A;Cross-references: EMBL:U07973; NID:9520454; PIDN:AAA83407.1; PID:9537432
 C;Genetics:
 A;Gene: COL3A1
 C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology:
 F;30-90/domain: von Willebrand factor type C repeat homology <VWC>

RESULT 21
 CGBOVS

collagen alpha 1(III) chain - bovine
 C;Species: Bos primigenius taurinus (cattle)
 C;Date: 04-Dec-1986 #sequence_revision 04-Dec-1986 #text_change 07-May-1999
 C;Accession: A02867; A38001; A38002; A38003; A38004; A38005; S71946
 R;Fietzek, P.; Altmann, H.; Rauterberg, J.; Heinkel, W.; Wachter, E.; Kuehn, K.
 Hoppe-Seyler's Z. Physiol. Chem. 360, 809-820, 1979
 A;Title: The covalent structure of calf skin type III collagen. I. The amino acid sequence
 A;Reference number: A02862; MUID:80026026
 A;Accession: A02862
 A;Molecule type: protein
 A;Residues: 1-242 <PHE>
 R;Deweis, H.; Fietzek, P.P.; Kuehn, K.
 Hoppe-Seyler's Z. Physiol. Chem. 360, 821-832, 1979
 A;Title: The covalent structure of calf skin type III collagen. II. The amino acid sequence
 A;Reference number: A38001; MUID:80026027
 A;Accession: A38001
 A;Molecule type: protein
 A;Residues: 243-422 <DENI>
 R;Benz, H.; Fietzek, P.P.; Kuehn, K.
 Hoppe-Seyler's Z. Physiol. Chem. 360, 833-840, 1979
 R;Lang, H.; Glanville, R.W.; Fietzek, P.P.; Kuehn, K.
 Hoppe-Seyler's Z. Physiol. Chem. 360, 841-850, 1979
 A;Title: The covalent structure of calf skin type III collagen. III. The amino acid sequence
 A;Reference number: A38002; MUID:80026028
 A;Accession: A38002
 A;Molecule type: protein
 A;Residues: 423-571 <BEN>
 R;Deves, H.; Lang, H.; Glanville, R.W.; Fietzek, P.P.; Kuehn, K.
 Hoppe-Seyler's Z. Physiol. Chem. 360, 851-860, 1979
 A;Title: The covalent structure of calf skin type III collagen. IV. The amino acid sequence
 A;Reference number: A38003; MUID:80026029
 A;Accession: A38003
 A;Molecule type: protein
 A;Residues: 552-808 <LAN>
 R;Deves, H.; Fietzek, P.P.; Kuehn, K.
 Hoppe-Seyler's Z. Physiol. Chem. 360, 861-868, 1979
 A;Title: The covalent structure of calf skin type III collagen. V. The amino acid sequence
 A;Reference number: A38004; MUID:80026030
 A;Accession: A38004
 A;Molecule type: protein
 A;Residues: 809-947 <DENW>
 R;Allmann, H.; Fietzek, P.P.; Kuehn, K.
 Hoppe-Seyler's Z. Physiol. Chem. 360, 869-876, 1979
 A;Title: The covalent structure of calf skin type III collagen. VI. The amino acid sequence
 A;Reference number: A38005; MUID:80026031
 A;Accession: A38005
 A;Molecule type: protein
 A;Residues: 948-1049 <ANL>
 A;Experimental source: skin
 R;Heinkel, W.
 Biochem. J. 318, 497-503, 1996
 A;Title: Cross-link analysis of the C-telopeptide domain from type III collagen
 A;Reference number: S71946; MUID:96404897
 A;Accession: S71946
 A;Molecule type: protein
 A;Residues: 87-106;1017-1029;1037-1049 <HEN>
 C;Comment: Prolines at the third position of the tripeptide repeating unit (G-X-Y) are
 C;Comment: The type III collagen molecule is a trimer of identical chains, linked to
 C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homologous
 C;Keywords: coiled coil; extracellular matrix; glycoprotein; hydroxylysine; hydroxyproline
 F;1-1049/Product: collagen alpha 1(III) chain #status experimental <CA>
 F;1-14/Region: amino-terminal nonhelical telopeptide

F; 587-589/Region: cell attachment (R-G-D) motif
 F; 752-754/Region: cell attachment (R-G-D) motif
 F; 875-877/Region: cell attachment (R-G-D) motif
 F; 880/Region: cell attachment (R-G-D) motif
 F; 935-937/Region: cell attachment (R-G-D) motif
 F; 950-terminal peptide: carboxyl-terminal nonhelical telopeptide
 F; 951-950/Modified site: allysine (Lys) #status predicted
 F; 107-binding site: carboxyl (Lys) (covalent) #status experimental
 F; 1040-1041/Disulfide bonds: interchain #status predicted

Query Match 65.2%; Score 60; DB 1; Length 1049;
 Best Local Similarity 71.4%; Pred. No. 0.54;
 Matches 10; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GVGDKGNPGWCA 14
 Db 702 GVGERSPGGCA 715

RESULT 22

CGU7L

Collagen alpha 1(III) chain precursor - human chain

N: Alternative names: procollagen alpha 1(III)

C: Species: Homo sapiens (man)

A: Cross-references: EMBL:X14420; NID:930057; PIDN:CAA32593.1; PID:930058

C: Accession: S05272; S04642; PE0011; S01726; S04887; A90399; A94562; I51868; S59511; A90399

R: Prockop, D.J.; submitted to the EMBL Data Library, February 1989

A: Submitted number: S05272

A: Accession: S05272

A: Status: preliminary

A: Molecule type: mRNA

A: Residues: 1-1240, 'V', 1242-1466 <PRC>

A: Cross-references: EMBL:X14420; NID:930057; PIDN:CAA32593.1; PID:930058

R: Alia-Kokko, L.; Kontusaari, S.; Baldwin, C.T.; Kulyanen, H.; Prockop, D.J.

A: Title: Structure of cDNA clones coding for the entire prepro-alpha1(III) chain of human collagen. J. 260, 509-516, 1989

A: Reference number: S0642; MUID:89350838

A: Accession: S0642

A: Molecule type: mRNA

A: Residues: 1-1196 <ALA>

A: Cross-references: EMBL:X14420; NID:930057; PIDN:CAA32593.1; PID:930058

A: Note: the complete sequence is not shown

R: Benson-Chanda, V.; Su, M.W.; Weil, D.; Chu, M.L.; Ramirez, F.

Gene 78, 255-265, 1989

A: Accession: S0642

A: Molecule type: mRNA

A: Residues: 1-1196 <ALA>

A: Cross-references: EMBL:X14420; NID:930057; PIDN:CAA32593.1; PID:930058

A: Note: the complete sequence is not shown

R: Benson-Chanda, V.; Su, M.W.; Weil, D.; Chu, M.L.; Ramirez, F.

Gene 78, 255-265, 1989

A: Title: Cloning and analysis of the 5' portion of the human type-III procollagen gene. C

A: Reference number: PEG011; MUID:89378752

A: Accession: PEG011

A: Molecule type: DNA

A: Residues: 1-176 <BENS>

A: Cross-references: GB:M2639; NID:9180813; PIDN:AAAS2040.1; PID:9180814

R: Toman, P.D.; Ricca, G.A.; de Crombrugge, B.

Nucleic Acids Res 16, 7201, 1988

A: Title: Nucleotide sequence of a cDNA coding for the amino-terminal region of human pre-pro-alpha1(III) collagen. J. Biol. Chem. 263, 5256-5259, 1988

A: Reference number: S01726; MUID:88303360

A: Molecule type: mRNA

A: Residues: 1-170 <TOM>

A: Cross-references: EMBL:X07240; NID:930060; PIDN:CAA30229.1; PID:930061

A: Note: the authors translated the codon CAG for residue 154 as HIS

R: Janecko, R.A.; Ramirez, F.

Nucleic Acids Res 17, 6742, 1989

A: Title: Nucleotide and amino acid sequences of the entire human alpha-1 (III) collagen. J. Biol. Chem. 265, 17070-17077, 1990

A: Cross-references: EMBL:X07240; NID:930060; PIDN:CAA30229.1; PID:930061

A: Note: the authors translated the codon CAG for residue 154 as HIS

R: Seyer, J.M.; Mainardi, C.; Kang, A.H.

Biochemistry 19, 1583-1589, 1980

A: Accession: S01726

A: Molecule type: mRNA

A: Residues: 1-170 <TOM>

A: Cross-references: EMBL:X07240; NID:930060; PIDN:CAA30229.1; PID:930061

A: Note: the authors translated the codon CAG for residue 154 as HIS

R: Janecko, R.A.; Ramirez, F.

Nucleic Acids Res 17, 6742, 1989

A: Title: Nucleotide and amino acid sequences of the entire human alpha-1 (III) collagen. J. Biol. Chem. 265, 17070-17077, 1990

A: Cross-references: EMBL:X07240; NID:930060; PIDN:CAA30229.1; PID:930061

A: Note: the authors translated the codon CAG for residue 154 as HIS

R: Seyer, J.M.; Kang, A.H.

Biochemistry 16, 1158-1164, 1977

A: Status: translation not shown

A: Title: Covalent structure of collagen: amino acid sequence of cyanogen bromide peptidase product. A90399; MUID:77134724

A: Reference number: A90399

A: Accession: A90399

A: Molecule type: protein

A: Residues: 'V', 169-225, 229-232, 'P', 234-292, 'D', 294-398 <SEY1>

A: Experimental source: liver

A: Note: sequence corrected by A94562; attachment of 2-O-alpha-D-glucosyl-O-beta-D-galactose to the Atlas, December 1977

R: Seyer, J.M.

Am. J. Hum. Genet. 53, 62-70, 1993

A: Reference number: A94562

A: Accession: A94562

A: Molecule type: protein

A: Residues: 'V', 169-225, 229-277, 'A', 279-292, 'D', 294, 'S', 296-398 <SEY2>

A: Experimental source: liver

A: Note: author submitted corrections to A90399

R: Milewicz, D.M.; Witz, A.M.; Smith, A.C.; Manchester, D.K.; Waldstein, G.; Byers, P.

A: Title: Parental somatic and germ-line mosaicism for a multiplexon deletion with unusual flanking sequence. A94562; MUID:93304430

A: Reference number: 151868; MUID:93304430

A: Status: preliminary; translated from GB/EMBL/DBJ

A: Molecule type: DNA

A: Cross-references: GB:S62225; NID:9386425; PIDN:AAB13937.1; PID:94261637

R: Chiodo, A.A.; Silience, D.O.; Cole, W.G.; Bateman, J.F.

Biochem. J. 311, 939-943, 1995

A: Title: Abnormal type III collagen produced by an exon-17-skipping mutation of the COL3A1 gene. A90414; MUID:97900343

A: Reference number: S59511; MUID:96067614

A: Accession: S59511

A: Molecule type: mRNA

A: Residues: 302-423 <CHI>

A: Cross-references: GB:S79877; NID:91195576; PIDN:AAB35615.1; PID:91195577

R: Seyer, J.M.; Kang, A.H.

Biochemistry 17, 3404-3411, 1978

A: Title: Covalent structure of collagen: amino acid sequence of five consecutive CNBr peptides. A90414

A: Residues: 399-675, 'N', 677-727 <SEY3>

A: Experimental source: liver

R: Lee, B.; Vitale, E.; Superti-Furga, A.; Steinmann, B.; Ramirez, F.

J. Biol. Chem. 266, 5256-5259, 1991

A: Title: G to T transversion at position +5 of a splice donor site causes skipping of the first intron. A90414; MUID:91161621

A: Reference number: I55349; MUID:91161621

A: Status: translated from GB/EMBL/DBJ

A: Molecule type: DNA

A: Residues: 537-605 <LEE>

A: Cross-references: GB:M59312; NID:9180815; PIDN:AAAS2041.1; PID:9180816

R: Seyer, J.M.; Mainardi, C.; Kang, A.H.

Biochemistry 19, 1583-1589, 1980

A: Title: Covalent structure of collagen: amino acid sequence of alpha1 (III)-CB5 from human skin fibroblasts. A90438; MUID:80198282

A: Reference number: A90438

A: Molecule type: protein

A: Residues: 728-895, 'A', 897-964 <SEY4>

A: Experimental source: liver

R: Cole, W.G.; Chioldo, A.A.; Lamande, S.R.; Janecko, R.; Ramirez, F.; Dahl, H.H.M.; Cole, W.G.; Chioldo, A.A.; Janecko, R.; Ramirez, F.

J. Biol. Chem. 265, 17070-17077, 1990

A: Title: A base substitution at a splice site in the COL3A1 gene causes exon skipping. A: Reference number: A38303

A: Molecule type: mRNA

A: Residues: 861-915 <COL>

A: Cross-references: GB:J05617; GB:M55603; GB:MS9227; NID:9180878; PIDN:AAB59383.1; PID:9180878

R: Mankoff, B.S.; Daigle, R.

Nucleic Acids Res. 16, 2337, 1988

A: Title: Human pro-alpha1(III) collagen: cDNA sequence for the 3' end. A: Reference number: S02119; MUID:88168927

A: Accession: S02119

A;Molecule type: mRNA
A;Residues: 950-1018; 'Y', 1020-1183, 'S', 1185-1466 <MAN>
A;Cross-references: EMBL:X06700; NID:g30053; PIDN:CAA29886_1; PID:g30054
R;Sever, J.M.; Kang, A.H.
Biochemistry 20, 2621-2627, 1981
A;Title: Covalent structure of collagen: amino acid sequence of alpha1 (III)-CB9 from type I
A;Reference number: A90446; MUID:81208139
A;Accession: A90446
A;Molecule type: protein
A;Residues: 965-979; 'A', 981-984, 'PS', 987, 'QN', 990-1095, 'P', 1098-1152, 'AT', 1155, 'S', 1157-
A;Title: Molecular cloning and carboxyl-propeptide analysis of human type III procollagen
A;Reference number: A93551; MUID:85087944
A;Molecule type: mRNA
A;Residues: 1155, 'P', 1157-1466 <LOI>
A;Cross-references: EMBL:X01655; EMBL:X01742; NID:g29884; PIDN:CAA25021_1
R;Miskulin, M.; Daigleish, R.; Klavne-Beckerhan, B.; Renard, S.I.; Tolstoshev, P.; Brant
Biochemistry 25, 1408-1413, 1986
A;Title: Human type III collagen gene expression is coordinately modulated with the type
A;Reference number: 152393; MUID:86187804
A;Accession: 152393
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1161-1200 <MTS>
A;Cross-references: GB:M13146; NID:g180415; PIDN:AAA52003_1; PID:g180416
R;Emanuel, B.S.; Cannizaro, L.A.; Sevier, J.M.; Myers, J.C.
Proc. Natl. Acad. Sci. U.S.A. 82, 3385-3389, 1985
A;Title: Human alpha 1(III) and alpha 2(V) procollagen genes are located on the long arm
A;Reference number: 159025; MUID:85216505
A;Accession: 1779359
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Cross-references: GB:M11134; NID:g180417; PIDN:AAA50004_1; PID:g180418
R;Chu, M.L.; Weil, D.; de Wet, W.; Bernard, M.; Sippola, M.; Ramirez, F.
J. Biol. Chem. 260, 4357-4363, 1985
A;Title: Isolation of cDNA and genomic clones encoding human pro-alpha1(III) collagen. Part
A;Reference number: A92516; MUID:85157600
A;Accession: A92516
A;Molecule type: DNA
A;Residues: 1176-1240, 'V', 1242-1356, 'P', 1358-1466 <CHI>
A;Cross-references: GB:M10615; GB:M10793; GB:M10794; GB:M10795; GB:M10796; GB:M10797; GE
A;Experimental source: liver
A;Title: the authors translated the codon TTC for residue 1057 as Tyr; the codons given in
A;Comment: Prolines and lysines at the third position of the tripeptide repeating unit
C;Comment: About 15% of the lysines are 5-hydroxylated and some are subsequently C
C;Genetics:
A;Gene: GDB:COL3A1
A;Cross-references: GDB:118729; OMIM:120180
A;Map position: 2q31-2q31
A;Introns: 27/1; 94/3; 111/3; 149/3; 176/3; 554/3; 587/3; 1175/3; 1275/1; 1337/3; 1418/3
A;Note: the list of introns is incomplete; defects in this gene can result in Ehlers-Danlos syndrome
C;Complex-type III collagen is a homotrimer of monomers initially linked by disulfide bonds of their length, is formed with desmosine cross-links made from lysine and allysine
C;Function:
A;Description: structural component of extracellular fibrous polymer that maintains integrity
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology; glycoprotein; hyd
C;Keywords: coiled coil; Ehlers-Danlos syndrome; extracellular matrix; glycoprotein; hyd
F;1-23/Domain: signal sequence #status predicted <SIG>
F;2-4/153/Domain: amino-terminal propeptide #status predicted <PRO>
F;31-91/Domain: von Willebrand factor type C repeat homology <WNC>
F;154-172/Domain: collagen alpha 1(III) chain #status predicted <MAT>
F;155-167/Region: amino-terminal nonhelical telopeptide
F;168-1196/Region: helical
F;1091-1093/Region: cell attachment (R-G-D) motif
F;1197-1221/Region: carboxyl-terminal nonhelical telopeptide
F;1222-1466/Domain: carboxyl-terminal propeptide #status predicted <CP>
F;1238-1466/Domain: fibrillar collagen carboxyl-terminal homology <FC>
F;24/Modified site: pyrrolidone carboxylic acid (Gin) (in mature form) #status predicted

F;153-154/Cleavage site: Pro-Gln (procollagen N-endopeptidase) #status predicted
F;154/Modified site: pyrrolidone carboxylic acid (Gin) (in mature form) #status pred
F;161-1212/Modified site: allysine (Lys) #status predicted
F;63-284, 850-97, 1106/Modified site: 5-hydroxylysine (Lys) #status experimental
F;263/Binding site: carbohydrate (Lys) (covalent) #status experimental
F;948-949/Cleavage site: Gly-Xle (collagenase) #status experimental
F;1106/Binding site: carbohydrate (Lys) (covalent) #status predicted
Query Match 65 2%; Score 60; DB 1; Length 1466;
Best Local Similarity 78.6%; Pred. No. 0.76; Pred. No. 0.76;
Matches 11; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 GVKGDKGNPGWPGA 14
Db 738 GPKGDKGEPGCPGA 751

RESULT 23
A61262
A;Cross-references: GB:M01669; NID:g179516; PIDN:AAA35605_1; PID:g179517
N;Alternate names: bullous pemphigoid 180K autoantigen BPAG2; bullous pemphigoid anti
C;Species: Homo sapiens (man)
C;Date: 12-May-1994 #sequence_revision 12-Jul-1996 #text_change 21-Jul-2000
C;Accession: 156325; 155345; A61262
R;Giudice, G.J.; Emory, D.J.; Diaz, L.A.
J. Invest. Dermatol. 99, 243-250, 1992.
A;Title: Cloning and primary structural analysis of the Bullous pemphigoid autoantige
A;Reference number: 156325; MUID:92381323
A;Accession: 156325
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1532 <RES>
A;Cross-references: GB:M01669; NID:g179516; PIDN:AAA35605_1; PID:g179517
R;Li, K.H.; Sawamura, D.; Giudice, G.J.; Diaz, L.A.; Mattei, M.G.; Chu, M.L.; Ultto,
J. Biol. Chem. 266, 24062-24069, 1991
A;Title: Genomic organization of collagenous domains and chromosomal assignment of hu
A;Reference number: 155345; MUID:92084712
A;Accession: 155345
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 543-890, 'P', <RE2>
A;Cross-references: GB:M03730; NID:g179520; PIDN:AAA51839_1; PID:g179521
R;Giudice, G.J.; Squierra, H.L.; Elias, P.M.; Diaz, L.A.
J. Clin. Invest. 87, 734-738, 1991
A;Title: Identification of two collagen domains within the bullous pemphigoid autoant
A;Reference number: A61262; MUID:9123476
A;Accession: A61262
A;Molecule type: mRNA
A;Residues: 543-890, 'P', <GIU>
C;Genetics:
A;Gene: GDB:COL1A1; BPAG2; BP180
A;Cross-references: GDB:131396; OMIM:113811
A;Map position: 10q24.3-10q24.3

Query Match 65.2%; Score 60; DB 2; Length 1532;
Best Local Similarity 78.6%; Pred. No. 0.79; Pred. No. 0.79;
Matches 11; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 GVKGDKGNPGWPGA 14
Db 995 GPKGDKGEPGCPGA 1008

RESULT 24
T13990
Collagen type IV alpha 2 - fruit fly (*Drosophila melanogaster*)
C;Species: *Drosophila melanogaster*
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 17-Nov-2000
C;Accession: T13990
R;Yasothornskri, S.; Davis, W.J.; Cramer, G.; Kimbrell, D.A.; Dearolf, C.R.
submitted to the EMBL Data Library, July 1995
A;Description: Viking: identification and characterization of a novel type IV collage

A: Reference number: 217845
A: Accession: TI390
A: Status: preliminary; translated from GB/EMBL/DDJB
A: Molecule type: mRNA
A: Residues: 1-1761 <XAS>
A: Cross-references: EMBL:U65431; NID:92281290; PID:92281291; PIDN:AAB64082.1
C: Superfamily: collagen alpha 1(IV) chain

Query Match 65.2%; Score 50; DB 2; Length 1761;
Best Local Similarity 66.7%; Pred. No. 0.91; 2; Mismatches 3; Indels 0; Gaps 0;
Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 1 GVKGDKGNPGWQCAP 15
Db 994 GRIGETENPGFGRP 1008

RESULT 25

A45407

Collagen alpha 3(IV) chain - sea urchin (*Strongylocentrotus purpuratus*)
C: Date: 22-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 13-Aug-1999
C: Accession: A45407; A45903; A25940
R: Exposito, J.Y.; D'Alessio, M.; Di Liberto, M.; Ramirez, F.
J. Biol. Chem. 268, 5249-5254, 1993
A: Title: Complete primary structure of a sea urchin type IV collagen alpha chain and analysis
A: Reference number: A45407; MUID:93186842
A: Accession: A45407
A: Status: preliminary; not compared with conceptual translation
A: Molecule type: nucleic acid
A: Residues: 1-1752 <EXP>
A: Note: sequence extracted from NCBI backbone (NCBIP:126841)
R: Wessel, G.M.; Etkin, M.; Benson, S.
Dev. Biol. 148, 261-272, 1991
A: Title: Primary mesenchyme cells of the sea urchin embryo require an autonomously produced
A: Reference number: A43903; MUID:92038439
A: Accession: A43903
A: Status: preliminary
A: Molecule type: mRNA
A: Residues: 'P', 633->537, 'G', <WNS>
A: Cross-references: GB:S64572; PIDN:AAE20270.1; PID:9238617
A: Note: sequence extracted from NCBI backbone (NCBIM:64572; NCBIP:64573)
R: Venkatesan, M.; De Pablo, F.; Vogeil, G.; Simpson, R.T.
Proc. Natl. Acad. Sci. U.S.A. 83, 3351-3355, 1986
A: Title: Structure and developmentally regulated expression of a *Strongylocentrotus purpuratus* gene
A: Reference number: A23940; MUID:86205894
A: Accession: A23940
A: Residues: 742-812 <ENV>
A: Molecule type: DNA
A: Cross-references: EMBL:M13206
C: Keywords: collagen alpha 1(IV) chain
F162-161-/Domain: amino-terminal nonhelical, 75 <7SD>
F162-1523/Region: interrupted helical
F1524-1752/Domain: carboxyl-terminal nonhelical, NCL <NCL>
F1534-1634/Domain: collagen IV carboxyl-terminal repeat <CT1>
F1644-1748/Domain: collagen IV carboxyl-terminal repeat <CT2>
F1719/Modified site: allysine (Tys) #status predicted

Query Match 64.7%; Score 59.5; DB 2; Length 1752;
Best Local Similarity 54.2%; Pred. No. 1.1; 2; Mismatches 13; Conservative 0; Mismatches 2; Indels 9; Gaps 1;

Qy 1 GVKGDKGNPGWQCAP 15
Db 1177 GVKGDPGRTGPEGAKGNGLPGIP 1200

RESULT 26

S33603

surfactant protein D - bovine
C: Species: Bos primigenius taurus (cattle)
C: Date: 02-Dec-1993 #sequence_revision 01-sep-1995 #text_change 17-Mar-1999
C: Accession: S33603
R: Lim, B.L.; Lu, J.; Reid, K.B.M.
Immunology 78, 159-165, 1993
A: Title: Structural similarity between bovine conglutinin and bovine lung surfactant
A: Reference number: S33603; MUID:93170856
A: Status: preliminary
A: Molecule type: mRNA
A: Residues: 1-369 <LIM>
C: Superfamily: pulmonary surfactant protein D; C-type lectin homology <LCH>
F; 248-367/Domain: C-type lectin homology <LCH>
Query Match 64.1%; Score 59; DB 2; Length 369;
Best Local Similarity 66.7%; Pred. No. 0.27; 2; Mismatches 3; Indels 0; Gaps 0;
Matches 10; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GVKGDKGNPGWQCAP 15
Db 157 GLKGERGAPGDPFGAP 171

RESULT 27

A39024

Collagen alpha 3(IV) chain - bovine (fragment)
C: Species: Bos primigenius tauris (cattle)
C: Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 13-Aug-1999
C: Accession: A39024; S20672; S17802; A35167; C39419; S17474; S20815
R: Morrison, K.E.; Germino, G.G.; Reeder, S.T.
J. Biol. Chem. 261, 7074-7077, 1987
A: Title: Use of the polymerase chain reaction to clone and sequence a cDNA encoding the alpha-3(IV) chain of basement membrane
A: Reference number: A39024; MUID:9103146
A: Accession: A39024
A: Molecule type: mRNA
A: Residues: 1-471 <MOR>
A: Cross-references: EMBL:M63139; NID:9162886; PIDN:AAA62708.1; PID:9162887
R: Butkowski, R.J.; Langenfeld, J.P.M.; Wieslander, J.; Hamilton, J.; Hudson, B.G.
J. Biol. Chem. 262, 7074-7077, 1987
A: Title: Localization of the Goodpasture epitope to a novel chain of basement membrane
A: Reference number: S18432; MUID:87222419
A: Accession: S18432
A: Molecule type: protein
A: Residues: 227-228, 'X', 230-244 <BUT>
R: Saus, J.; Wieslander, J.; Langenfeld, J.P.M.; Quinones, S.; Hudson, B.G.
J. Biol. Chem. 263, 1374-1380, 1988
A: Title: Identification of the Goodpasture antigen as the alpha-3(IV) chain of collagen
A: Reference number: S17802; MUID:88330844
A: Accession: S17802
A: Molecule type: protein
A: Residues: 227-228, 'X', 230-252, 'Y', 254 <SAU>
R: Gunwar, S.; Saus, J.; Noelken, M.E.; Hudson, B.G.
J. Biol. Chem. 265, 5466-5469, 1990
A: Title: Glomerular basement membrane. Identification of a fourth chain, alpha4, of the basement membrane
A: Reference number: A35167; MUID:90202779
A: Accession: A35167
A: Molecule type: protein
A: Residues: 236-258 <GUN>
R: Gunwar, S.; Ballesteros, F.; Kalluri, R.; Timoneda, J.; Chonko, A.M.; Edwards, S.J.; J. Biol. Chem. 266, 15118-15124, 1991
A: Title: Glomerular basement membrane. Identification of dimeric subunits of the noncollagenous domain
A: Reference number: A39419; MUID:91332055
A: Accession: C39419
A: Molecule type: protein
A: Residues: 236-255 <GUN>
C: Superfamily: collagen alpha 1(IV) chain
C: Keywords: basement membrane; cell binding; coiled coil; disulfide bond; duplication
F1-238/Domain: collagenous (fibrillar) #status predicted <COL>
F1-239-471/Domain: carboxyl-terminal nonhelical, NCL #status predicted <NCL>
F1-239-473/Domain: repeat NCL #status predicted <NCL>
F1-354-471/Domain: repeat NCL #status predicted <NCL>
F; 232, 238/Modified site: hydroxyproline (Pro) #status experimental

Fri Nov 1 13:09:20 2002

us-09-529-691a-1.rpr

Page 17

Job time : 15.5 secs

THIS PAGE BLANK (uspto)

OM protein - protein search, using sw model

Run on: November 1, 2002, 12:16:06 ; Search time 7 Seconds
 (without alignments)
 82.970 Million cell updates/sec

Title:	US-09-529-691A-1
Scoring table:	BLOSSUM62
Sequence:	1 GVKGDKGNGWPGAP 15
Perfect score:	92
Score:	92
Scanned:	Total number of hits satisfying chosen parameters: 10524
	Minimum DB seq length: 0
	Maximum DB seq length: 200000000
Post-processing:	Maximum Match 0%
Database :	SwissProt_40;*
	Listing first 45 summaries
	SUMMARIES
Result	Query
No.	Match Length DB ID
1	92 100.0 1669 1 CA14_HUMAN
2	78 84.8 1669 1 CA14_MOUSE
3	69 75.0 1685 1 CAS4_HUMAN
4	71.7 754 1 CA54_CANFA
5	65 70.7 1707 1 CA24_MOUSE
6	64 69.6 1712 1 CA24_HUMAN
7	63 68.5 2944 1 CA17_HUMAN
8	62 67.4 1670 1 CA34_HUMAN
9	61 66.3 339 1 COAS5_BPPRD
10	61 66.3 1603 1 CA1F_HUMAN
11	61 66.3 1690 1 CA44_HUMAN
12	61 66.3 1763 1 CA24_ASCSU
13	60 65.2 369 1 PSPD_BOVIN
14	60 65.2 623 1 CA44_RABBIT
15	60 65.2 636 1 CA13_RAT
16	60 65.2 1049 1 CA13_BOVIN
17	60 1262 1210 1 CA13_CHICK
18	60 65.2 1466 1 CA13_HUMAN
19	59 64.1 471 1 CA34_BOVIN
20	59 64.1 1758 1 CA14_CAEEL
21	58 63.0 285 1 YRF2_CAEEL
22	58 63.0 1516 1 CA1H_HUMAN
23	58 63.0 1758 1 CA24_CAEEL
24	57 62.0 298 1 CC34_CAEEL
25	57 62.0 483 1 MRCO_MESAU
26	57 62.0 1356 1 CA2L_ONCMY
27	56 60.9 518 1 MTCO_MOUSE
28	56 60.9 1143 1 CA1H_HUMAN
29	56 60.9 1464 1 CA13_MOUSE
30	55.5 60.3 458 1 MSRE_MOUSE
31	55 59.8 245 1 CLOC_HUMAN
32	55 59.8 266 1 YWKW_CAEEL
33	55 59.8 371 1 CONG_BOVIN

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

ALIGNMENTS

RESULT ¹	CAL1_HUMAN	STANDARD:	PRT: 1669 AA.
ID	CA14_HUMAN		
AC	P02462	(Rel. 01, Created)	
DT	21-JUL-1986	(Rel. 33, last sequence update)	
DT	01-OCT-1996	(Rel. 40, Last annotation update)	
DE	Collagen alpha 1(IV), chain precursor.		
GN	COLA1L		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE-89340433; PubMed=2701944;		
RA	Soininen R., Huotari M., Ganguly A., Prockop D.J., Tryggvason K.;		
RT	"Structural organization of the gene for the alpha 1 chain of human type IV collagen";		
RT	type IV collagen";		
RL	J. Biol. Chem. 264:13565-13571(1989).		
RN	[2]		
RP	SEQUENCE OF 46-1257 FROM N.A.		
RC	TISSUE=Placenta;		
RX	MEDLINE-88003384; PubMed=3691802;		
RA	Soininen R., Haka-Risku T., Prockop D.J., Tryggvason K.;		
RT	"Complete primary structure of the alpha 1-chain of human basement membrane (type IV) collagen";		
RT	FEBS Lett. 225:188-194(1987).		
RL	[3]		
RN			
RP	SEQUENCE OF 1-943 FROM N.A.		
RC	TISSUE=Placenta;		
RX	MEDLINE-88029471; PubMed=3311751;		
RA	Brazel D., Oberbaumer I., Bieringer H., Babel W., Gianville R.W., Deutzmann R., Kuehn K.;		
RT	"Completion of the amino acid sequence of the alpha 1 chain of human basement membrane collagen (type IV) reveals 21 non-triplet interruptions located within the collagenous domain";		
RT	Interruptions located within the collagenous domain";		
RT	Eur. J. Biochem. 168:529-536(1987).		
RL	[4]		
RN			
RP	SEQUENCE OF 28-243.		
RX	MEDLINE-860040108; PubMed=4043082;		
RA	Gianville R.W., Qian R.Q., Sibley B., Ristell J., Kuehn K.;		
RT	"Amino acid sequence of the N-terminal aggregation and cross-linking region (TS domain) of the alpha 1 (IV) chain of human basement membrane collagen";		
RT	membrane collagen";		
RT	membrane collagen";		
RL	Eur. J. Biochem. 152:213-219(1985).		
RN	[5]		
RP	SEQUENCE OF 534-1447.		
RX	MEDLINE-85033629; PubMed=6434307;		
RA	Babel W., Gianville R.W.;		
RT	"Structure of human-basement-membrane (type IV) collagen. Complete amino-acid sequence of a 914-residue-long pepsin fragment from the alpha 1(IV) chain";		
RT	alpha 1(IV) chain";		
RL	Eur. J. Biochem. 143:545-556(1984).		
RN	[6]		
RP	SEQUENCE OF 1256-1669 FROM N.A.		

RX MEDLINE=95207819; PubMed=2581969;
 RA Pihlajaniemi T., Tryggvason K., Myers J.C., Kurkinen M., Lebo R.,
 Cheung M.-C., Prockop D.J., Boyd C.D.;
 RT "cDNA clones coding for the pro-alpha1(IV) chain of human type IV
 procollagen reveal an unusual homology of amino acid sequences in two
 halves of the carboxyl-terminal domain";
 RT J. Biol. Chem. 260:7681-7687(1985).
 RL RN [7]
 RP SEQUENCE OF 1259-1669 FROM N.A.
 RX MEDLINE=95216555; PubMed=258242;
 RA Brinker J.M., Gudas L.J., Loidl H.R., Wang S.-Y., Rosenblum J.,
 Refalides N.A., Myers J.C.;
 RA "Restricted homology between human alpha 1 type IV and other
 procollagen chains";
 RT PROC. NATL. Acad. Sci. U.S.A. 82:3649-3653(1985).
 RN [8]
 RP SEQUENCE OF 1-28 FROM N.A.
 RX MEDLINE=95034231; PubMed=3182844;
 RA Sozin R., Huotari M., Hostikkka S.L., Prockop D.J., Tryggvason K.;
 RT "The structural genes for alpha 1 and alpha 2 chains of human type IV
 collagen are divergently encoded on opposite DNA strands and have an
 overlapping promoter region";
 RT J. Biol. Chem. 263:17217-17220(1988).
 RN [9]
 RP SEQUENCE OF 141-1669, AND DISULFIDE BONDS.
 TISSUE=Placenta;
 RC MEDLINE=9505112; PubMed=2844531;
 RA Siebold B., Deutmann R., Kuhn K.;
 RT "The arrangement of intra- and intermolecular disulfide bonds in the
 carboxyterminal, non-collagenous aggregation and cross-linking domain
 of basement-membrane type IV collagen";
 RT Eur. J. Biochem. 176:617-624(1988).
 CC GLOMERULAR BASEMENT MEMBRANES (GBM), FORMING A 'CHICKEN-WIRE'
 CC MESHWORD TOGETHER WITH LAMININS, PROTEOGLYCANS AND ENACTIN/
 CC NIDOGEN.
 CC -!- SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOFORMS, ALPHA 1(IV)-
 CC ALPHA 6(IV), EACH OF WHICH CAN FORM A TRIPLE HELIX STRUCTURE
 CC WITH 2 OTHER CHAINS TO GENERATE TYPE IV COLLAGEN NETWORK.
 -!- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS
 DOMAIN (NCL) AT THEIR C-TERMINUS; FREQUENT INTERRUPTIONS OF THE
 G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY
 CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL
 TRIPLE-HELICAL 7S DOMAIN.
 CC -!- PTM: LYSINES AT THE THIRD POSITION OF THE TRIPETIDE REPEATING
 CC UNIT ARE HYDROXYLATED IN ALL CASES AND BIND CARBONATE.
 -!- DOMAIN: PROLINES AT THE THIRD POSITION OF THE TRIPETIDE REPEATING
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 -!- PTM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH
 CC ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF
 CC THESE, LOCATED IN THE NCL DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE
 CC IV COLLAGENS.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC DR EMBL: M26548; AAA51098.1; JOINED.
 CC DR EMBL: M26549; AAA51098.1; JOINED.
 CC DR EMBL: M26551; AAA51098.1; JOINED.
 CC DR EMBL: M26552; AAA51098.1; JOINED.
 CC DR EMBL: M26553; AAA51098.1; JOINED.
 CC DR EMBL: M26554; AAA51098.1; JOINED.
 CC DR EMBL: M26555; AAA51098.1; JOINED.
 CC DR EMBL: M26556; AAA51098.1; JOINED.
 CC DR EMBL: M26557; AAA51098.1; JOINED.
 CC DR EMBL: M26539; AAA51098.1; JOINED.
 CC DR EMBL: M26558; AAA51098.1; JOINED.
 CC DR EMBL: M26559; AAA51098.1; JOINED.
 CC DR EMBL: M26560; AAA51098.1; JOINED.
 CC DR EMBL: M26561; AAA51098.1; JOINED.
 CC DR EMBL: M26562; AAA51098.1; JOINED.
 CC DR EMBL: M26536; AAA51098.1; JOINED.
 CC DR EMBL: M26563; AAA51098.1; JOINED.
 CC DR EMBL: M26568; AAA51098.1; JOINED.
 CC DR EMBL: M26541; AAA51098.1; JOINED.
 CC DR EMBL: M26564; AAA51098.1; JOINED.
 CC DR EMBL: M26565; AAA51098.1; JOINED.
 CC DR EMBL: M26566; AAA51098.1; JOINED.
 CC DR EMBL: M26567; AAA51098.1; JOINED.
 CC DR EMBL: M26574; AAA51098.1; JOINED.
 CC DR EMBL: M26575; AAA51098.1; JOINED.
 CC DR EMBL: M26569; AAA51098.1; JOINED.
 CC DR EMBL: M26570; AAA51098.1; JOINED.
 CC DR EMBL: M26571; AAA51098.1; JOINED.
 CC DR EMBL: M26572; AAA51098.1; JOINED.
 CC DR EMBL: M26566; AAA51098.1; JOINED.
 CC DR EMBL: M26573; AAA51098.1; JOINED.
 CC DR EMBL: M26574; AAA51098.1; JOINED.
 CC DR EMBL: M26575; AAA51098.1; JOINED.
 CC DR EMBL: M26561; CA29075.1; --.
 CC DR EMBL: M10340; AAA52006.1; --.
 CC DR EMBL: M11315; AAA52042.1; --.
 CC DR PIR; S16976; CGHU4B.
 CC DR PIR; 120130; --.
 CC DR InterPro; IPR001442; C4.
 CC DR InterPro; IPR00087; Collagen.
 CC DR Pfam; PF01413; C4; 2.
 CC DR Pfam; PF01391; Collagen; 21.
 CC DR SMART; SM0011; C4; 2.
 CC DR KW Extracellular matrix; Connective tissue; Basement membrane;
 CC DR Repeat; Hydroxylation; Glycoprotein; Collagen; Signal.
 CC FT SIGNAL 1 27
 CC FT PROPEP 28 172
 CC FT CHAIN 173 1669
 CC FT DOMAIN 173 1440
 CC FT DOMAIN 1441 1669
 CC FT CARBOYD 126 126
 CC FT DISULFID 1460 1551
 CC FT DISULFID 1493 1548
 CC FT DISULFID 1505 1511
 CC FT DISULFID 1570 1665
 CC FT DISULFID 1604 1662
 CC FT DISULFID 1616 1622
 CC FT DISULFID 237 238
 CC FT CONFLICT 241 241
 CC FT CONFLICT 319 319
 CC FT CONFLICT 719 719
 CC FT CONFLICT 837 837
 CC FT CONFLICT 842 842
 CC FT CONFLICT 895 896
 CC FT CONFLICT 904 904
 CC FT CONFLICT 914 914
 CC FT CONFLICT 998 998
 CC FT CONFLICT 1010 1010
 CC FT CONFLICT 1012 1012
 CC FT CONFLICT 1358 1358
 CC FT CONFLICT 1669 1669
 CC SO AA; 160611 MW.; 3BEBAA6DFFB9B8A84 CRC64;

Query Match 100 %; Score 92; DB 1; Length 1669;
 Best Local Similarity 100.0%; Pred. No. 5.3e-06;

Matches 15; conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 2

Oy 1 GVKGDKGNPWGCGAP 15
 ID CA14_MOUSE STANDARD; PRT; 1669 AA.
 AC P02463;
 DT 01-FEB-1991 (Rel. 01, created)
 DT 21-JUL-1986 (Rel. 17, last sequence update)
 DE 15-JUL-1999 (Rel. 38, Last annotation update)
 RT Collagen alpha 1(IV) chain precursor.

GN COL4A1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. NEBI_TAXID=10900;
 RN [1]

RP SEQUENCE FROM N.A. MEDLINE=89197932; PubMed=2703490;

RA Muthukumaran G., Blumberg B., Kurkinen M.; "the complete primary structure for the alpha 1-chain of mouse collagen IV. Differential evolution of collagen IV domains.", J. Biol. Chem. 264:6310-6317(1989).

[2]

RP SEQUENCE OF 1-1154 FROM N.A. MEDLINE=88112221; PubMed=3338568;

RX Wood L., Theriault N., Vogeli G.; "cDNA clones completing the nucleotide and derived amino acid sequence of the alpha 1 chain of basement membrane (type IV) collagen from mouse.", FEBS Lett. 227:5-8(1988).

RN [3]

RP SEQUENCE OF 1149-1424 FROM N.A. MEDLINE=86301886; PubMed=3755692;

RA Nath P., Laurent M., Horn E., Sobel M.E., Zon G., Vogeli G.; "Isolation of an alpha 1 type-IV collagen cDNA clone using a synthetic oligodeoxynucleotide.", Gene 43:301-304(1986).

RN [4]

RP SEQUENCE OF 1276-1669 FROM N.A. MEDLINE=8512033; PubMed=2379861;

RA Oberbaumer T., Laurent M., Schwarz U., Sakurai Y., Yamada Y., Vogeli G., Voss T., Siebold B., Glanzville R.W., Kuhn K.; "Amino acid sequence of the non-collagenous globular domain (NC1) of the alpha 1(IV) chain of basement membrane collagen as derived from complementary DNA.", Eur. J. Biochem. 147:217-224(1985).

[5]

RP SEQUENCE OF 1441-1669 FROM N.A. MEDLINE=87230460; PubMed=3597383;

RA Kurkinen M., Condon M.R., Blumberg B., Barlow D., Quinones S., Saus J., Pihlajaniemi T.; "Extensive homology between the carboxyl-terminal peptides of mouse alpha 1(IV) and alpha 2(IV) collagen.", J. Biol. Chem. 262:8496-8499(1987).

RN [6]

RP PARTIAL SEQUENCE FROM N.A. MEDLINE=86156099; PubMed=3009468;

RA Sakurai Y., Sullivan M., Yamada Y.; "Alpha 1 type IV collagen gene evolved differently from fibrillar collagen genes.", J. Biol. Chem. 261:6654-6657(1986).

RN [7]

RP SEQUENCE OF 1-28 FROM N.A. MEDLINE=89056738; PubMed=3198626;

RA Keytes P., Wood L., Theriault N., Kurkinen M., Vogeli G.; "Head-to-head arrangement of murine type IV collagen genes.", J. Biol. Chem. 263:19274-19277(1988).

RN [8]

RP SEQUENCE OF 1-129 FROM N.A. MEDLINE=89071759; PubMed=3200851;

RA Burbelo P.D., Martin G.R., Yamada Y.; "Alpha 1(IV) and alpha 2(IV) collagen genes are regulated by a bidirectional promoter and a shared enhancer."; RT PROC. Natl. Acad. Sci. U.S.A. 85:9579-9582(1988); RL [9]

RN SEQUENCE OF 1-129 FROM N.A. MEDLINE=88243724; PubMed=3379041;

RA Killen P.D., Burbelo P., Sakurai Y., Yamada Y.; "Structure of the amino-terminal portion of the murine alpha 1(IV) collagen chain and the corresponding region of the gene.", J. Biol. Chem. 263:8706-8709(1988).

-1 FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF GLOMERULAR BASEMENT MEMBRANES (GBM), FORMING A 'CHICKEN-WIRE' MESHWORK TOGETHER WITH LAMININS, PROTEOGLYCANS AND ENACTIN/ NITROGEN.

CC -1 SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOFORMS, ALPHA 1(IV)-ALPHA 6(IV), EACH OF WHICH CAN FORM A TRIPLE HELIX STRUCTURE WITH 2 OTHER CHAINS TO GENERATE TYPE IV COLLAGEN NETWORK.

CC -1 DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL TRIPLE-HELICAL 7S DOMAIN.

CC -1 PTM: PROLINES AT THE THIRD POSITION OF THE TRIPETIDE REPEATING UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.

CC -1 PTM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH ARE INVOLVED IN INNER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF THESE, LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE IV COLLAGENS.

CC -----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/>) or send an email to license@isb-sib.ch).

CC -----

DR EMBL; J03758; AAA37439.;
 DR EMBL; M23333; AAA51625.;
 DR EMBL; J04694; AAA50292.;
 DR EMBL; X06777; CAM29946.;
 DR EMBL; X02201; CAM26132.;
 DR EMBL; M15832; AAA37340.;
 DR EMBL; M14042; AAA37342.;
 DR EMBL; M12879; AAA37343.;
 DR EMBL; M13024; --; NOT_ANNOTATED_CDS.
 DR EMBL; M13025; --; NOT_ANNOTATED_CDS.
 DR EMBL; M13026; AAA37344.;
 DR EMBL; M13027; AAA37345.;
 DR EMBL; M13043; AAA37346.;
 DR EMBL; J04448; AAA37437.;
 DR PIR; A33525; CGMSB.
 DR MGD; M88154; Col4al.
 DR InterPro; IPR00142; C4.
 DR InterPro; IPR000087; Collagen.
 DR Pfam; PF01413; C4; 2.
 DR Pfam; PF01391; Collagen; 21.
 DR PRODOM; PDD03923; C4; 2.
 DR SMART; SM00111; C4; 2.
 KW Extracellular matrix; Connective tissue; Basement membrane; Repeat; Hydroxylation; Signal.

FT SIGNAL 1
 FT PROPEP 28 172 AMINO-TERMINAL PROPEPTIDE (7S DOMAIN).

FT CHAIN 173 1669 COLLAGEN ALPHA 1(IV) CHAIN.
 FT DOMAIN 173 1440 TRIPLE-HELICAL REGION.
 FT DOMAIN 1441 1669 NONHELICAL REGION (NC1).

FT DISULFID 1460 1551 OR 1548 (BY SIMILARITY).
 FT DISULFID 1493 1548 OR 1551 (BY SIMILARITY).
 FT DISULFID 1505 1511 BY SIMILARITY.

FT DISULFID 1570 1665 OR 1662 (BY SIMILARITY).

FT DISULFID 1604 1662 OR 1665 (BY SIMILARITY).
 FT DISULFID 1616 1622 BY SIMILARITY.
 FT CARBOHYD 126 126 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 26 26 A -> P (IN REF. 2).
 FT CONFLICT 186 186 S -> L (IN REF. 2).
 FT CONFLICT 319 319 Q -> S (IN REF. 2).
 FT CONFLICT 369 369 Q -> L (IN REF. 2).
 FT CONFLICT 403 403 L -> F (IN REF. 2).
 FT CONFLICT 481 481 P -> L (IN REF. 2).
 FT CONFLICT 493 493 Q -> H (IN REF. 2).
 FT CONFLICT 712 712 S -> I (IN REF. 2).
 FT CONFLICT 813 813 E -> Q (IN REF. 2).
 FT CONFLICT 982 982 Q -> H (IN REF. 2).
 FT CONFLICT 1397 1397 V -> S (IN REF. 3).
 FT SEQUENCE 1669 AA; 16680 MW; 42916891E52058E9 CRC64;

Query Match 84 8%; Score 78; DB 1; Length 1669;
 Best Local Similarity 86.7%; Pred. No. 0.00065; Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GVKGDKGNPGWPGAP 15
 Db 1263 GPKGDKGNGQWPGAP 1277

RESULT 3

CAS5_HUMAN STANDARD; PRT: 1685 AA.

ID P29400; 016126; 01006; DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-FEB-1994 (Rel. 28, last sequence update)
 DT 01-MAR-2002 (Rel. 41, last annotation update)

DE Collagen alpha 5(IV) chain precursor.
 GN COL4A5.

OS Homo sapiens (Human).

OC Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

OX NCBI_TAXID=9606;
 RN [1] SEQUENCE FROM N.A.

RX MEDLINE=94165049; PubMed=812014;
 RA Zhou J., Leitonen A., Tryggvason K.;
 RT "Structure of the human type IV collagen COL4A5 gene.";
 RL J. Biol. Chem. 269:6608-6614(1994).

RX TISSUE=Kidney;
 RA MEDLINE=92316923; PubMed=1352287;
 RT Zhou J., Hertz J.M., Leitonen A., Tryggvason K.;
 RT "Complete amino acid sequence of the human alpha 5 (IV) collagen chain and identification of a single-base mutation in exon 23 converting glycine 521 in the collagenous domain to cysteine in an Alport syndrome patient.";
 RT J. Biol. Chem. 267:12475-12481(1992).

RN [3] SEQUENCE OF 85-1685 FROM N.A.
 RX TISSUE=Placenta;
 RA MEDLINE=00337990; PubMed=2380186;
 RA Pihlajaniemi T., Pohjolaainen E.R., Myers J.C.;
 RT "Complete primary structure of the triple-helical region and the carboxyl-terminal domain of a new type IV collagen chain, alpha 5(IV).";
 RT J. Biol. Chem. 265:13758-13766(1990).

RX SEQUENCE OF 924-1685 FROM N.A.
 RA MEDLINE=91169491; PubMed=2004755;
 RA Zhou J., Hostikka S.L., Chow L.T., Tryggvason K.;
 RT "Characterization of half of the human type IV collagen alpha 5 gene that is affected in the Alport syndrome.";
 RT Genomics 9:1-9(1991).

RN [5] SEQUENCE OF 914-1685 FROM N.A.
 RA MEDLINE=90160375; PubMed=1689491;

RA Hostikka S.L., Eddy R.L., Byers M.G., Hoeyhvae M., Shows T.B., Tryggvason K.;
 RA "Identification of a distinct type IV collagen alpha chain with restricted kidney distribution and assignment of its gene to the locus of X chromosome-linked Alport syndrome.";
 RT Proc. Natl. Acad. Sci. U.S.A. 87:1606-1610(1990).
 RN [6] SEQUENCE OF 1442-1471 FROM N.A.
 RX MEDLINE=90252791; PubMed=2339699;
 RA Myers J.C., Jones T.A., Pohjolaainen E.R., Kadri A.S., Goddard A.D., Sheer D., Solomon E., Pihlajaniemi T.;
 RT "Molecular cloning of the X chromosome containing the gene to the region of the X chromosome containing the Alport syndrome locus.";
 RT Ann. J. Hum. Genet. 46:1024-1033(1990).
 RN [7] SEQUENCE OF 1-20 FROM N.A.
 RA Guo C., van Damme B., Vanrenterghem Y., Devriendt K., Cassiman J.-J., Marynen P.;
 RA Submitted (SEP-1994) to the EMBL/GenBank/DDJB databases.
 RL [8] SEQUENCE OF 1258-1270 FROM N.A. (SPLICED FORM).
 RX MEDLINE=94133540; PubMed=8301933;
 RA Guo C., van Damme B., van Damme-Lombaerts R., van den Berghe H., Cassiman J.-J., Marynen P.;
 RT "Differential splicing of COL4A5 mRNA in kidney and white blood cells: a complex mutation in the COL4A5 gene of an Alport patient.";
 RA Kidney Int. 44:1316-1321(1993).
 RL [9] "The clinical spectrum of type IV collagen mutations.";
 RT Lemmink H.H., Schroeder C.H., Monnens L.A.H., Smeets H.J.M.;
 RL Hum. Mutat. 9:477-499(1997).
 RN [10] REVIEW ON VARIANTs.
 RX MEDLINE=9738662; PubMed=9195222;
 RA Zhou J., Barker D.F., Hostikka S.L., Gregory M.C., Atkins C.L., Tryggvason K.;
 RT "Single base mutation in alpha 5(IV) collagen chain gene converting a conserved cysteine to serine in Alport syndrome.";
 RT Genomics 9:10-18(1991).
 RN [11] VARIANT AS SER-1564.
 RX MEDLINE=91169492; PubMed=1672282;
 RA Zhou J., Barker D.F., Hostikka S.L., Gregory M.C., Atkins C.L., Tryggvason K.;
 RT "Substitution of arginine for glycine 325 in the collagen alpha 5 (IV) chain associated with X-linked Alport syndrome: characterization of the mutation by direct sequencing of PCR-amplified lymphoblast cDNA fragments.";
 RT Ann. J. Hum. Genet. 51:135-142(1992).
 RL [12] VARIANT AS GLU-325.
 RX MEDLINE=92303539; PubMed=1376965;
 RA Knebelmann B., Deschenes G., Gros F., Hors M.-C., Gruenfeld J.-P., Tryggvason K., Gubler M.-C., Antignac C.;
 RT "Substitution of arginine for glycine 325 in the collagen alpha 5 (IV) chain associated with X-linked Alport syndrome: characterization of the mutation by direct sequencing of PCR-amplified lymphoblast cDNA fragments.";
 RT Ann. J. Hum. Genet. 51:135-142(1992).
 RL [13] VARIANT AS GLU-325.
 RX MEDLINE=93244772; PubMed=1363780;
 RA Ronieri A., Seri M., Myers J.C., Pihlajaniemi T., Massella L., Rizzoni G.F., de Marchi M.;
 RT "De novo mutation in the COL4A5 gene converting glycine 325 to glutamic acid in Alport syndrome.";
 RT Hum. Mol. Genet. 1:127-129(1992).
 RN [13] VARIANT AS THR-1517; SER-1538 AND GLN-1563.
 RX MEDLINE=9401048; PubMed=8400498;
 RA Lemmink H.H., Schroeder C.H., Brunner H.G., Nelen M.R., Zhou J., Tryggvason K., Haggema-Schouten W.A.G., Roordaets A.P., Rascher W., van Oost B.A., Smeets H.J.M.;
 RT "Identification of four novel mutations in the COL4A5 gene of patients with Alport syndrome.";
 RL Genomics 17:485-489(1993).
 RN [14] VARIANT AS E-400; V-406; V-638; A-638; R-653; R-796; R-869; R-872
 RP AND C-1241.

RX MEDLINE=95322976; PubMed=7599631;
 RA Boye E., Flintner F., Zhou J., Tryggvason K., Borrows M., Harris A.;
 RT "detection of 12 novel mutations in the collagenous domain of the
 COL4A5 gene in Alport syndrome patients.";
 RT Hum. Mutat. 5:197-204(1995).
 RN [15]
 RP VARIANT AS ARG-1649.
 RX MEDLINE=96213750; PubMed=86551292;
 RA Barker D.F., Pruchno C.J., Jiang X., Atkin C.L., Stone E.M.,
 Denison J.C., Fair P.R., Gregory M.C.
 RT "A mutation causing Alport syndrome with tardive hearing loss in
 common in the western United States";
 RL Am. J. Hum. Genet. 58:1157-1165(1996).
 RN [16]
 RP VARIANT AS.
 RX MEDLINE=96213754; PubMed=86551296;
 RA Renieri A., Bruttini M., Galli L., Zanelli P., Neri T.M., Rossetti S.,
 RA Scalari F., Sesia A., Rizzoni G.F., Tryggvason K., Pignatti P.F.,
 RA Savi M., Ballabio A., de Marchi M.;
 RT "X-linked Alport syndrome: an SSCP-based mutation survey over all 51
 exons of the COL4A5 gene.";
 RL Am. J. Hum. Genet. 58:1192-1204(1996).
 RN [17]
 RP VARIANT AS, AND VARIANT AS-P-430; SER-444; SER-619; ASN-664 AND
 RP MET-1428.
 RX MEDLINE=97094179; PubMed=8940267;
 RA Knebelmann B., Breillat C., Forestier L., Arrendel C., Jaccassier D.,
 RA Giatras I., Drouot L., Deschenes G., Grunfeld J.-P., Broyer M.,
 RA Gubler M.-C., Antignac C.;
 RT "Spectrum of mutations in the COL4A5 collagen gene in X-linked Alport
 syndrome.";
 RL Am. J. Hum. Genet. 59:1221-1232(1996).
 RN [18]
 RP VARIANT AS ASP-1498.
 RX MEDLINE=96233932; PubMed=8829632;
 RA Tverskaya S., Bobryntseva V., Tsalykova F., Ignatova M.,
 RA Krasnopol'skaya X., Evrakov O.;
 RT "Common ancestry of three Ashkenazi-American families with Alport
 chain associated with adult-onset X-linked Alport syndrome.";
 RL Hum. Mutat. 7:149-150(1996).
 RN [19]
 RP VARIANT AS GLN-1677.
 RX MEDLINE=97295089; PubMed=9150741;
 RA Barker D.F., Denison J.C., Atkin C.L., Gregory M.C.;
 RT "Common ancestry of three Ashkenazi-American families with Alport
 syndrome and COL4A5 R167Q.";
 RL Hum. Genet. 99:681-684(1997).
 RN [20]
 RP VARIANT AS R-174; R-177; R-325; C-1410; W-1421; T-1517 AND D-1596.
 RX MEDLINE=98112435; PubMed=9152056;
 RA Neri T.M., Zanelli P., de Palma G., Savi M., Rossetti S., Turco A.E.,
 RA Pignatti G.F., Galli L., Bruttini M., Renieri A., Mangarelli R.,
 RA Trivellini A., Pinciaroli A.R., Reggiani M., Rizzoni G.F., de Marchi M.;
 RT "Missense mutations in the COL4A5 gene in patients with X-linked
 Alport syndrome.";
 RL Hum. Mutat. Suppl. 1:S106-S109(1998).
 RN [21]
 RP VARIANT AS V-420: 456-P--P-458 DEL; D-573; D-624; D-635; 802-G--P-807
 DEL; C-941; S-1030; S-1056; D-1143; R-1196; E-1261; S-1357
 RP AND R-1649.
 RX MEDLINE=99063529; PubMed=98488783;
 RA Martin P., Heiskanen N., Zhou J., Leinonen A., Tumelius T., Hertz J.M.,
 RA Barker D.F., Gregory M.C., Atkin C.L., Styrkardottir U., Neumann H.,
 RA Springer J., Shows T.B., Pettersson B., Tryggvason K.;
 RT "High mutation detection rate in the COL4A5 collagen gene in suspected
 RT Airport syndrome using PCR and direct DNA sequencing.";
 RL J. Am. Soc. Nephrol. 9:2291-2301(1998).
 RN [22]
 RP VARIANT AS GLU-579; LYS-623; ASN-947; VAL-953; ARG-1107; ARG-1158;
 RP SER-170 AND TRP-167, AND VARIANT SER-444 AND ALA-739.
 RX MEDLINE=20030197; PubMed=10561141;
 RA Inoue Y., Nishio H., Shirakawa T., Nakaniishi K., Nakamura H.,

RX MEDLINE=20025011; PubMed=10563487;
 RX Query Match 75.0%; Score 69; DB 1; Length 1685;
 RT Best Local Similarity 73.3%; Pred. No. 0.014;
 RT Matches 11; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 RN [23]
 RP VARIANT AS ARG-822.
 RX MEDLINE=20025011; PubMed=10563487;
 Rx RESULT 4
 RC CAS4_CANFA
 ID CAS4_CANFA
 STANDARD: PRT; 754 AA.
 RP Q28247;
 RT DT 01-NOV-1997 (Rel. 35, Created)
 DE DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE Collagen alpha 5(IV) chain (Fragment).
 GN COL4A5.
 OS Canis familiaris (Dog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi;
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
 OC NCBI_TAXID=9615;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-SAMOT; TISSUE-Kidney;
 RX MEDLINE=94224868; PubMed=8171024;
 RA Zheng K., Thornton P.S., Baurnal R., McInnes R.R.;
 RT "Canine X chromosome-linked hereditary nephritis: a genetic model for
 human X-linked hereditary nephritis resulting from a single base
 mutation in the gene encoding the alpha 5 chain of collagen type
 IV.";
 RT Proc. Natl. Acad. Sci. U.S.A. 91:3989-3993(1994).
 RL -1- FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF
 CC GLOMERULAR BASEMENT MEMBRANES (GBM), FORMING A 'CHICKEN-WIRE'
 CC MESHWORK TOGETHER WITH LAMININS, PROTEOGLYCANS AND ENACTIN/
 CC NIDGEN.
 CC -1- SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOFORMS, ALPHA 1(IV)-
 CC ALPHA 6(IV), EACH OF WHICH CAN FORM A TRIPLE HELIX STRUCTURE
 CC WITH 2 OTHER CHAINS TO GENERATE TYPE IV COLLAGEN NETWORK.
 CC -1- SUBCELLULAR LOCATION: CELL SURFACE (POTENTIAL).
 CC -1- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS
 CC DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE
 CC G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE HELICAL DOMAIN (WHICH MAY
 CC CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL
 CC TRIPLE-HELICAL 7S DOMAIN.
 CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPETIDE REPEATING
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 CC -1- PTM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH
 CC ARE INVOLVED IN INNER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF
 CC THESE, LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE
 CC IV COLLAGENS.
 CC -1- DISEASE: A DEFECT IN COL4A5 HAS BEEN FOUND TO BE THE CAUSE OF
 CC CANINE X-LINKED HEREDITARY NEPHRITIS (HN), A DISEASE SIMILAR TO
 CC THAT IN HUMANS (ALSO REFERRED TO AS ALPORT SYNDROME) CHARACTERIZED
 CC BY PROGRESSIVE RENAL FAILURE AND NEUROSENSORY DEAFNESS.
 CC -1- SIMILARITY: TO OTHER TYPE IV COLLAGENS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>).
 CC or send an email to license@isb-sib.ch).

DR	InterPro; IPR001442; C4.	RT	"Nucleotide sequence coding for the human type IV collagen alpha 2 chain cDNA reveals extensive homology with the NC-1 domain of alpha 1 (IV) but not with the collagenous domain or 3'-untranslated region.";
PFAM	PF01413; C4; 2.	RT	"The complete primary structure of the alpha 2 chain of human type IV collagen and comparison with the alpha 1(IV) chain.";
PRODOM	PD003923; C4; 2.	RL	"Human basement membrane collagen (type IV). The amino acid sequence of the alpha 2(IV) chain and its comparison with the alpha 1(IV) chain reveals deletions in the alpha 1(IV) chain.";
SMART	SM00111; C4; 2.	RN	Eur. J. Biochem. 172:35-42(1988).
DR	Extracellular matrix; Connective tissue; Repeat; Hydroxylation; Glycoprotein; Basement membrane; Collagen; Signal.	RX	MEDLINE=8702438; PubMed=3035878;
KW	SIGNAL	RA	Grieffin C.A., Emanuel B.S., Hansen J.R., Cavenee W.K., Myers J.C.;
FT	25	RT	"Human collagen genes encoding basement membrane alpha 1(IV) and alpha 2 (IV) chains map to the distal long arm of chromosome 13.";
PROPEP	25	RA	Myers J.C., Howard P.S., Jelen A.M., Dion A.S., Macarak E.J.;
FT	183	RT	"Duplication of type IV collagen COOH-terminal repeats and species-specific expression of alpha 1(IV) and alpha 2(IV) collagen genes.";
CHAIN	184	RL	Proc. Natl. Acad. Sci. U.S.A. 84:512-516(1987).
FT	1707	RN	[5]
DOMAIN	184	FT	SEQUENCE OF 1451-1485 FROM N.A.
FT	1479	FT	MEDLINE=87250571; PubMed=2439508;
DISULFID	1499	FT	Myers J.C., Howard P.S., Jelen A.M., Dion A.S., Macarak E.J.;
FT	1588	FT	"Duplication of type IV collagen COOH-terminal repeats and species-specific expression of alpha 1(IV) and alpha 2(IV) collagen genes.";
FT	1585	FT	J. Biol. Chem. 262:9231-9238(1987).
FT	1588	FT	[6]
FT	1588	FT	SEQUENCE OF 1-33 FROM N.A.
FT	1588	FT	MEDLINE=89054231; PubMed=3182844;
FT	1588	FT	Spoininen R., Huotari M., Hostikka S.L., Prockop D.J., Tryggvason K.;
FT	1588	FT	"The structural genes for alpha 1 and alpha 2 chains of human type IV collagen are divergently encoded on opposite DNA strands and have an overlapping promoter region.";
FT	1588	FT	J. Biol. Chem. 263:17217-17220(1988).
FT	1588	FT	[7]
FT	1588	FT	SEQUENCE OF 1-33 FROM N.A.
FT	1588	FT	MEDLINE=89030632; PubMed=2846280;
FT	1588	FT	Poeschl E., Poliner R., Kuehn K.;
FT	1588	FT	"The genes for the alpha 1(IV) and alpha 2(IV) chains of human basement membrane collagen type IV are arranged head-to-head and separated by a bidirectional promoter of unique structure.";
FT	1588	FT	EMBO J. 7:2687-2695(1988).
FT	1588	FT	[8]
SEQUENCE	1653	FT	SEQUENCE OF 1-33 FROM N.A.
FT	1653	FT	RX
CARBOHYD	138	FT	MEDLINE=93305049; PubMed=8317999;
FT	1270	FT	Fischer G., Schmidt C., Optiz J., Cully Z., Kuehn K., Poeschl E.;
FT	1270	FT	"Identification of a novel sequence element in the common promoter region of human collagen type IV genes, involved in the regulation of divergent transcription.";
FT	1051	FT	RT
CONFLICT	1051	FT	"The structural genes for alpha 1 and alpha 2 chains of human type IV collagen are divergently encoded on opposite DNA strands and have an overlapping promoter region.";
FT	1097	FT	J. Biol. Chem. 263:17217-17220(1988).
FT	1097	FT	[7]
FT	1171	FT	SEQUENCE OF 1-33 FROM N.A.
FT	1171	FT	MEDLINE=89030632; PubMed=2846280;
FT	1179	FT	Poeschl E., Poliner R., Kuehn K.;
FT	1179	FT	"The genes for the alpha 1(IV) and alpha 2(IV) chains of human basement membrane collagen type IV are arranged head-to-head and separated by a bidirectional promoter of unique structure.";
FT	1241	FT	EMBO J. 7:2687-2695(1988).
FT	1241	FT	[8]
FT	1328	FT	SEQUENCE OF 1-33 FROM N.A.
FT	1328	FT	MEDLINE=89030632; PubMed=2846280;
FT	1573	FT	Poeschl E., Poliner R., Kuehn K.;
FT	1573	FT	"The genes for the alpha 1(IV) and alpha 2(IV) chains of human basement membrane collagen type IV are arranged head-to-head and separated by a bidirectional promoter of unique structure.";
FT	1623	FT	J. Biol. Chem. 263:17217-17220(1988).
SEQUENCE	1707	FT	[9]
QY	1	FT	SEQUENCE OF 1-33 FROM N.A.
Db	1165	FT	RX
1165	GIPGDKGDFGPGCP	FT	MEDLINE=89054231; PubMed=3182844;
1165	1179	FT	Spoininen R., Huotari M., Hostikka S.L., Prockop D.J., Tryggvason K.;
RESULT	6	FT	"The structural genes for alpha 1 and alpha 2 chains of human type IV collagen are divergently encoded on opposite DNA strands and have an overlapping promoter region.";
CA24_HUMAN	STANDARD;	FT	J. Biol. Chem. 263:17217-17220(1988).
ID	CA24_HUMAN	FT	[7]
AC	P08572;	FT	SEQUENCE OF 1-33 FROM N.A.
DR	01-AUG-1988 (Rel. 08, Created)	FT	MEDLINE=89030632; PubMed=2846280;
DT	01-FEB-1991 (Rel. 17, Last sequence update)	FT	Poeschl E., Poliner R., Kuehn K.;
DT	16-OCT-2001 (Rel. 40, Last annotation update)	FT	"The arrangement of intra- and intermolecular disulfide bonds in the carboxy-terminal, non-collagenous aggregation and cross-linking domain of basement-membrane type IV collagen.";
DE	Collagen alpha 2(IV) chain precursor.	FT	Eur. J. Biochem. 176:617-624(1988).
GN	COLA42.	FT	"FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF GLYCEROLIC BASEMENT MEMBRANES (GBM), FORMING A 'CHICKEN-WIRE' MESHWORK TOGETHER WITH LAMININS, PROTEOGLYCANS AND ENTACTIN/ NIDGENE."
OS	Homo sapiens (Human).	FT	CC
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	FT	-1 SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOFORMS, ALPHA 1(IV)-ALPHA 6(IV), EACH OF WHICH CAN FORM A TRIPLE HELIX STRUCTURE WITH 2 OTHER CHAINS TO GENERATE TYPE IV COLLAGEN NETWORK.
OX	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	FT	-1 DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS DOMAIN (NC1) AT THEIR C-TERMINUS. FREQUENT INTERRUPTIONS OF THE G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL TRIPLE-HELICAL 7S DOMAIN.
RN	[1]	FT	-1 PTM: ROLINES AT THE THIRD POSITION OF THE TRIPÉPTIDE REPEATING UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
RN	SEQUENCE OF 1-1042 FROM N.A.	FT	-1 PTM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF THESE LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE IV COLLAGENS.
RN	SEQUENCE-PLACEA;	FT	CC
RN	MEDLINE=88151998; PubMed=3345760;	FT	CC
RN	Brazel D., Pollner R., Oberhaeuser I., Kuehn K.;	FT	CC
RN	"Human basement membrane collagen (type IV). The amino acid sequence of the alpha 2(IV) chain and its comparison with the alpha 1(IV) chain reveals deletions in the alpha 1(IV) chain.";	FT	CC
RN	Eur. J. Biochem. 172:35-42(1988).	FT	CC
RN	[3]	FT	CC
RN	SEQUENCE OF 1-1712 FROM N.A.	FT	CC
RN	HOSTIKKA S.L., Kirkkinen M., Tryggvason K.;	FT	CC
RN	MEDLINE=87250571; PubMed=2439508;	FT	CC
RN	Hostikka S.L., Kirkkinen M., Tryggvason K.;	FT	CC

the European Bioinformatics Institute. There are no restrictions on its use by non profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/>) or send an email to license@isb-sib.ch).

CC EMBL; X05562; CAA29076; 1. -.

CC EMBL; X05610; CAA29098; 1. -.

CC EMBL; J02760; AAA58422; 1. -.

CC EMBL; M36963; AAA52099; 1. -.

CC EMBL; X12784; CAA31275; 1. -.

CC EMBL; J04217; AAA53097; 1. -.

CC PIR; A32024; A32024.

DR MIM; 120890; -.

DR InterPro; IPR001442; C4.

DR Pfam; PF01413; C4; 2.

DR Pfam; PF01391; Collagen; 20.

DR ProDom; PDO003913; C4; 2.

DR SMART; SM00111; C4; 2.

KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;

KW Glycoprotein; Basement membrane; Collagen; Signal.

FT PROPEP 26 183 AMINO-TERMINAL PROPEPTIDE (7S DOMAIN).

FT CHAIN 184 1712 COLLAGEN ALPHA 2(IV) CHAIN.

FT DOMAIN 184 1484 TRIPLE-HELICAL REGION.

FT DISULFID 1485 1712 NONHELCICAL REGION (NCL).

FT DISULFID 1504 1593 OR 1590 (BY SIMILARITY).

FT DISULFID 1537 1590 OR 1593 (BY SIMILARITY).

FT DISULFID 1549 1555 BY SIMILARITY.

FT DISULFID 1612 1708 OR 1705 (BY SIMILARITY).

FT DISULFID 1646 1655 OR 1708 (BY SIMILARITY).

FT CARBOYD 1658 1665 BY SIMILARITY.

FT CONFLICT 138 138 N-LINKED (GLCNAC. . .).

FT CONFLICT 471 683 R -> P (IN REF. 2).

FT CONFLICT 683 683 A -> G (IN REF. 2).

FT CONFLICT 1575 1575 M -> T (IN REF. 5).

FT CONFLICT 1663 1663 G -> H (IN REF. 9).

FT CONFLICT 1701 1701 H -> G (IN REF. 9).

FT SEQUENCE 1712 AA; 167535 MW; 2582A17847890037 CRC64;

Query Match, 69.6%; Score 64; DB 1; Length 1712;

Best Local Similarity 73.3%; Pred. No. 0.081; 2; Mismatches 2; Indels 0; Gaps 0; Oligo 1 GVKGDKGNPGWPGAP 15

Db 1368 GPKGKRGDPGFPGPAP 1382

RESULT 7

CAT1.HUMAN ID -CA17.HUMAN STANDARD; PRT; 2944 AA.

AC Q02388; Q14054; Q16507; DT 01-JUN-1994 (Rel. 29, Created)

DT 01-FEB-1996 (Rel. 33, Last sequence update)

DT 01-MAR-2002 (Rel. 41, Last annotation update)

DE Collagen alpha 1(VII) chain precursor (long-chain collagen) (LC DE Collagen).

GN COL7A1.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

OX NCBI_TAXID=9606; [1]

RN SEQUENCE FROM N.A.

RX MEDLINE=94327588; PubMed=8051117;

RX Christianson A.M., Greenspan D.S., Lee S., Utton J.;

RX "Cloning of human type VII collagen. Complete primary sequence of the alpha 1(VII) chain and identification of intragenic polymorphisms.";

RX J. Biol. Chem. 269:20256-20262(1994).

[2] SEQUENCE OF 128-1493 FROM N.A., AND PARTIAL SEQUENCE.

RX MEDLINE=93338417; PubMed=1307247;

RX Christiano A.M., Rosenbaum L.M., Chung-Honet L.C., Parente M.G., Woodley D.T., Pan T.C., Zhang R.Z., Chu M.-L., Burgeson R.E., Utton J.;

RX "The large non-collagenous domain (NC-1) of type VII collagen is amino-terminal and chimeric. Homology to cartilage matrix protein, the type III domains of fibronectin and the A domains of von Willebrand factor."; Curr Opin Mol. Genet. 1:475-481(1992).

RN [3] SEQUENCE OF 815-1439 FROM N.A.

RX MEDLINE=9134380; PubMed=1871109;

RX Parente M.G., Chung L.C., Ryynanen J., Woodley D.T., Wynn K.W., Bauer E.A., Mattei M.-G., Chu M.-L., Utton J.;

RX "Human type VII collagen: cDNA cloning and chromosomal mapping of the gene"; Proc. Natl. Acad. Sci. U.S.A. 88:6931-6935(1991).

RN [4] SEQUENCE OF 365-1255 FROM N.A.

RX MEDLINE=93107742; PubMed=1469284;

RX Gammon W.R., Aernettey M.L., Padilla K.M., Prisayanh P.S., Cook M.E., Wright J., Briggaman R.A., Hunt S.W. III;

RX "Noncollagenous (NCL) domain of collagen VII resembles multidomain adhesion proteins involved in tissue-specific organization of extracellular matrix"; J. Invest. Dermatol. 99:691-696(1992).

RN [5] SEQUENCE OF 340-675 FROM N.A.

RX TISSUE=keratinocytes;

RX MEDLINE=92231902; PubMed=1567409;

RX Tanaka T., Takahashi K., Furukawa F., Imamura S.;

RX "Molecular cloning and characterization of type VII collagen cDNA. A molecular cloning and characterization of type VII collagen cDNA"; R. Bloch. Biophys. Res. Commun. 183:958-963(1992).

RN [6] SEQUENCE OF 2395-2944 FROM N.A.

RX MEDLINE=93271985; PubMed=8499016;

RX Greenspan D.S.;

RX "The carboxyl-terminal half of type VII collagen, including the non-collagenous NC-2 domain and intron/exon organization of the corresponding region of the COL7A1 gene.";

RX Hum. Mol. Genet. 2:273-278(1993).

RN [7] SEQUENCE OF 1-87 FROM N.A.

RX TISSUE=placenta;

RX MEDLINE=94375010; PubMed=8088784;

RX Christiano A.M., Hoffman G.G., Chung-Honet L.C., Lee S., Cheng W., Utton J., Greenspan D.S.;

RX "Structural organization of the human type VII collagen gene (COL7A1), composed of more exons than any previously characterized gene.";

RX Genomics 21:169-179(1994).

RN [8] REVIEW ON DEB VARIANTS.

RX MEDLINE=9404166; PubMed=9375848;

RX Jaerikallio A., Puukkinen L., Utton J.;

RX "Molecular basis of dystrophic epidermolysis bullosa: mutations in the type VII collagen (COL7A1)."; Hum. Mutat. 10:338-347(1997).

RN [9] VARIANT RDBB LYS-2798.

RX MEDLINE=92291877; PubMed=851326;

RX Christianson A.M., Greenspan D.S., Hoffman G.G., Zhang X., Tamai Y., Lin A.N., Dietz H.C., Honvanian A., Utton J.;

RX "A missense mutation in type VII collagen in two affected siblings with recessive dystrophic epidermolysis bullosa.";

RX Nat. Genet. 4:62-66(1993).

RN [10] VARIANT DDBB SR-2040.

RX MEDLINE=94224777; PubMed=8170945;

RX Christianson A.M., Ryynanen M., Utton J.;

RX "Dominant dystrophic epidermolysis bullosa: identification of a GLY->SER substitution in the triple-helical domain of type VII collagen.";

RX Proc. Natl. Acad. Sci. U.S.A. 91:3549-3553(1994).

[11] VARIANT PEB-DDEB CYS-2623.
RN
RP
RX MEDLINE=96081220; PubMed=0541842;
RA Christiano A.M., Lee J.Y.Y., Chen W.J., Laforgia S., Utton J.;
RT "Preterminal epidermolysis bullosa: genetic linkage to COL7A1 and
RT identification of a glycine-to-cysteine substitution in the triple-
helical domain of type VII collagen.";
RL Hum. Mol. Genet. 4:1579-1583(1995).
[12] VARIANT DDEB ARG-2043.
RN
RP
RX MEDLINE=95164985; PubMed=7861014;
RA Christiano A.M., Morritone A., Paradisi M., Angelo C., Mazzanti C.,
CA Cavalieri R., Utton J.;
RT "A glycine-to-arginine substitution in the triple-helical domain of
RT type VII collagen in a family with dominant dystrophic epidermolysis
RT bullosa.";
RL J. Invest. Dermatol. 104:438-440(1995).
[13] VARIANT RDEB AND DDEB.
RN
RP
RX MEDLINE=96220218; PubMed=8644729;
RA Christiano A.M., McGrath J.A., Tan K.C., Utton J.;
RT "Glycine substitutions in the triple-helical region of type VII
RT collagen result in a spectrum of dystrophic epidermolysis bullosa
RT phenotypes and patterns of inheritance.";
RL Am. J. Hum. Genet. 58:671-681(1996).
RN [14] VARIANT RDEB ARG-2575.
RN
RP
RX MEDLINE=96154068; PubMed=8592061;
RA Shimizu H., McGrath J.A., Christiano A.M., Nishikawa T., Utton J.;
RT "Molecular basis of recessive dystrophic epidermolysis bullosa:
RT genotype/phenotype correlation in a case of moderate clinical
RT severity.";
RL J. Invest. Dermatol. 106:119-124(1996).
RN [15] VARIANT RDEB ARG-1782.
RN
RP
RX MEDLINE=96183562; PubMed=8618018;
RA Christiano A.M., McGrath J.A., Utton J.;
RT "Influence of the second COL7A1 mutation in determining the
RT phenotypic severity of recessive dystrophic epidermolysis bullosa.";
RL J. Invest. Dermatol. 106:766-770(1996).
RN [16] VARIANT RDEB ASP-2073.
RN
RP
RX MEDLINE=96310789; PubMed=8757759;
RA Dunnill M.G.S., McGrath J.A., Richards A.J., Christiano A.M.,
RA Utton J., Pope R.A.J.;
RT "Clinicopathological correlations of compound heterozygous COL7A1
RT mutations in recessive dystrophic epidermolysis bullosa.";
RL J. Invest. Dermatol. 107:171-177(1996).
RN [17] VARIANT RDEB W-1982; G-2008; A-2025; E-2049; G-2063 AND
RP R-2575.
RX MEDLINE=97465605; PubMed=8326325;
RA Horvath A., Rochat A., Bodemer C., Petit E., Rivers C.A., Prost C.,
RA Fraiture S., Christiano A.M., Utton J., Lthrop M., Barrandon Y.,
RA de Prost Y.;
RT "Characterization of 18 new mutations in COL7A1 in recessive
RT dystrophic epidermolysis bullosa provides evidence for distinct
RT molecular mechanisms underlying anchoring fibril
RT formation.";
RL Am. J. Hum. Genet. 61:599-610(1997).
RN [18] VARIANT RDEB ARG-1652.
RN
RP
RX MEDLINE=98106792; PubMed=9444387;
RA Cserhalmi-Friedman P.B., Karpati S., Horvath A., Christiano A.M.;
RT "Identification of a glycine substitution and a splice site mutation
RT in the type VII collagen gene in a proband with mitis recessive
RT dystrophic epidermolysis bullosa.";
RL Arch. Dermatol. Res. 289:640-645(1997).
RN [19] VARIANT DEB ARG-2009 AND ARG-2043.
RX MEDLINE=9738588; PubMed=9315684;
RA Winberg J.-O., Hammarlin-Hauslin N., Nilssen O., Anton-Lamprecht I.,
RA Naylor S.L., Kerbacher K., Zimmermann M., Krajci P.,
RA

RA Gedde-Dahl T. Jr., Bruckner-Tuderman L.;
RT "Modulation of disease severity of dystrophic epidermolysis bullosa by
RT a splice site mutation in combination with a missense mutation in the
COL7A1 gene.";
RT Hum. Mol. Genet. 6:1125-1135(1997).
RN [20] VARIANT DDEB ASP-1519; ASP-2006; GLU-2015 AND ARG-2034.
RN
RP
RX MEDLINE=98334662; PubMed=9668111;
RA Hamamli-Hauslin N., Schumann H., Raghunath M., Kilgus O., Luethi U.,
RA Lueger T., Bruckner-Tuderman L.;
RT "Some, but not all, glycine substitution mutations in COL7A1 result in
RT intracellular accumulation of collagen VII, loss of anchoring
RT fibrils, and skin blistering.";
RL J. Biol. Chem. 273:19228-19234(1998).
RN [21] VARIANT DEB CYS-2008; ARG-2207 AND SER-2775.
RN
RP
RX MEDLINE=98410969; PubMed=9740253;
RA Kon A., Pulkkinen L., Ishida-Yamamoto A., Hashimoto I., Utton J.;
RT "Novel COL7A1 mutations in dystrophic forms of epidermolysis
RT bullosa.";
RL J. Invest. Dermatol. 111:534-537(1998).
RN [22] VARIANT RDEB ARG-1347.
RN
RP
RX MEDLINE=99019477; PubMed=9804332;
RA Terracina M., Postoraro P., Schubert M., Sonnega G., Atzori F.,
RA Zambruno G., Bruckner-Tuderman L., Castiglione D.;
RT "Compound heterozygosity for recessive glycine substitution and a
RT splice site mutation in the COL7A1 gene causes an unusually mild form
RT of localized recessive dystrophic epidermolysis bullosa.";
RL J. Invest. Dermatol. 111:744-750(1998).
RN [23] VARIANT DEB TRP-2034; VAL-2040; ARG-2043; ARG-2064 AND ASP-2713.
RN MEDLINE=99072663; PubMed=9856843;
RX

Query	Match	Best Local Similarity	Score	DB	Length
QY	1	GVKGDKGNPGWGAP	68.5%	1	2944
Db	2386	GVKGDLGLPGLPGAP	2400		

RESULT 8
CA34_HUMAN STANDARD; PRM: 1670 AA.
ID CA34_HUMAN
AC 001955;
DT 01-OCT-1996 (Rel. 34, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Collagen alpha 3(IV) chain precursor (Goodpasture antigen).
GN COL4A3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OK NCBI_TAXID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RX MEDLINE=94344949; PubMed=8083201;
RA Maruyama M., Leinonen A., Mochizuki T., Tryggvason K., Reeder S.T.;
RT "Complete primary structure of the human alpha 3(IV) collagen chain.
RT Coexpression of the alpha 3(IV) and alpha 4(IV) collagen chains in
RT human tissues.";
RL J. Biol. Chem. 269:23013-23017(1994).
RN [2]
RP REVISIONS.
RA Leinonen A.;
RL Submitted (Oct-1998) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP SEQUENCE OF 1386-1670 FRM N.A. AND PARTIAL SEQUENCE.
RX MEDLINE=93015826; PubMed=100291;
RA Quinones S., Bernal D., Garcia-Sogo M., Elena S.F., Saus J.;

RT *Exon/intron structure of the human alpha 3(IV) gene encompassing the
 RT Goodpasture antigen (alpha 3(IV)NC1). Identification of a potentially
 RT antigenic region at the triple helix/NC1 domain junction.;"
 RL J. Biol. Chem. 267:19780-19784 (1992).

RN [4]

RP SEQUENCE OF 1453-1670 FROM N.A.
 RP MEDLINE=11353570; PubMed=1883840;
 RA RA "Sequence and localization of a partial cDNA encoding the human alpha
 RT 3 chain of type IV collagen.";
 RL Am. J. Hum. Genet. 49:545-554 (1991).

RN [5]

RP SEQUENCE OF 1331-1570 FROM N.A.
 RC TISSUE="Kidney";
 RX MEDLINE=93147878; PubMed=1737849;
 RA RA Turner N., Mason P.J., Brown R., Fox M., Povey S., Rees A.,
 RA Pusey C.D.; "Molecular cloning of the human Goodpasture antigen demonstrates it
 to be the alpha 3 chain of type IV collagen.";
 RL J. Clin. Invest. 89:592-601 (1992).

RN [6]

RP SEQUENCE OF 1544-1670 FROM N.A.
 RC TISSUE="Kidney";
 RA Ding J.;
 RL Submitted (JAN-1993) to the EMBL/GenBank/DDBJ databases.
 RN [7]

RP SEQUENCE OF 1439-1670, AND ALTERNATIVE SPlicing.
 RC TISSUE="Kidney";
 RX MEDLINE=94124597; PubMed=8294492;
 RA Feng L., Xia Y., Wilson C.B.;
 RT "Alternative splicing of the NC1 domain of the human alpha 3(IV)
 collagen gene. Differential expression of mRNA transcripts that
 predict three protein variants with distinct carboxyl regions.";
 RL J. Biol. Chem. 269:2342-2348 (1994).

RP SEQUENCE OF 1-29 FROM N.A.
 RX MEDLINE=88196354; PubMed=9537506;
 RA Momota R., Sugimoto M., Oohashi T., Kigasawa K., Yoshioka H.,
 RA Ninomiya Y.;
 RT "Two genes, COL4A3 and COL4A4 coding for the human alpha3(IV) and
 alpha4(IV) collagen chains are arranged head-to-head on chromosome
 2936.";
 RL FEBS Lett. 424:11-16 (1998).

RN [9]

RP ALTERNATIVE SPlicing.
 RX MEDLINE=93280184; PubMed=8505332;
 RA Bernal D., Quinones S., Saus J.;
 RT "The human mRNA encoding the Goodpasture antigen is alternatively
 spliced.";
 RL J. Biol. Chem. 268:12090-12094 (1993).

RN [10]

RP VARIANT PRO-174.
 RX MEDLINE=95078827; PubMed=7987301;
 RA Lemmink H.H., Mochizuki T., van den Heuvel L.P.W.J., Schroeder C.H.,
 RA Reenders S.T., Monnens L.A.H., van Oost B.A., Brunner H.G.,
 RA Bartolets A., Monnens L.J.M.,
 RL "Mutations in the type IV collagen alpha 3 (COL4A3) gene in autosomal
 recessive Alport syndrome.";
 RL Hum. Mol. Genet. 3:1269-1273 (1994).

RN [11]

RP VARIANTS AS E-297; R-407; R-1167; E-1207; Q-1215; S-1277; T-
 RP 1330; E-1334; E-1347; AND C-161, AND VARIANTS R-43; E-162; Y-326; H-
 RP 408; R-451; I-574; E-1269 AND P-1474.
 RX MEDLINE=21064696; PubMed=11142255;
 RA Helicot L., Arrendo C., Forestier L., Cohen-Solal L., Mollet G.,
 RA Gutierrez B., Stavrou C., Gubler M.C., Antignac C.,
 RT "Structure of the human type IV collagen gene COL4A3 and mutations in
 autosomal Alport syndrome.";
 RL J. Am. Soc. Nephrol. 12:97-106 (2001).

J. -i- FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF
 CC GLOMERULAR BASEMENT MEMBRANES (GBM), FORMING A 'CHICKEN-WIRE'
 CC MESHWORK TOGETHER WITH LAMININS, PROTEOGLYCANS AND ENACTIN/
 CC NIDOCIN.

CC -!- SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOFORMS, ALPHA 1(IV)-
 CC ALPHA 6(IV), EACH OF WHICH CAN FORM A TRIPLE HELIX STRUCTURE
 CC WITH 2 OTHER CHAINS TO GENERATE TYPE IV COLLAGEN NETWORK.
 CC -!- SUBCELLULAR LOCATION: CELL SURFACE (POTENTIAL).
 CC -!- ALTERNATIVE PRODUCTS: AT LEAST 3 ISOFORMS; 1 (SHOWN HERE), 2^u/ AND
 CC 3^u/LS; ARE PRODUCED BY ALTERNATIVE SPlicing. THEY DIFFER IN THEIR
 CC C-TERMINAL NC1 DOMAINS.
 CC -!- TISSUE SPECIFICITY: ALPHA 3 AND ALPHA 4 TYPE IV COLLAGENS ARE
 CC COLOCALIZED AND PRESENT ONLY IN BASEMENT MEMBRANES OF KIDNEY, EYE,
 CC COCHLEA, LUNG AND BRAIN.
 CC -!- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS
 CC DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE
 CC G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY
 CC CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL
 CC TRIPLE-HELICAL 7S DOMAIN.
 CC -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPIDE REPEATING
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 CC DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE
 CC G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY
 CC CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL
 CC TRIPLE-HELICAL 7S DOMAIN.
 CC -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPIDE REPEATING
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 CC DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE
 CC G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY
 CC CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL
 CC TRIPLE-HELICAL 7S DOMAIN.
 CC -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPIDE REPEATING
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 CC DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE
 CC G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY
 CC CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL
 CC TRIPLE-HELICAL 7S DOMAIN.
 CC -!- DISEASE: ANTIBODIES AGAINST THE NC1 DOMAIN OF ALPHA3(IV) MEDIATE
 CC THE HUMAN AUTOIMMUNE DISEASE, GOODPASTURE SYNDROME, WHICH IS
 CC CHARACTERIZED BY HEMATURIA AND PULMONARY HEMORRHAGE.
 CC -!- DISEASE: DEFECTS IN COL4A3 ARE ASSOCIATED WITH THE TYPE I
 CC AUTOSOMAL RECESSIVE FORM OF ALPORT SYNDROME, AN HEREDITARY
 CC GLomerulonephropathy characterized by progressive renal failure,
 CC hematuria and deafness. The recessive form occurs equally between
 CC MALES AND FEMALES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC - the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.ebi-sib.ch/announce/>
 CC or send an email to license@ebi-sib.ch).
 CC -----
 DR EMBL: X80031; CAA56335_1;
 DR EMBL: M9293; AAA2610_1; -
 DR EMBL: S55790; AAB19637_1; -
 DR EMBL: M81379; AAA51556_1; -
 DR EMBL: L08650; AA52044_1; -
 DR EMBL: U02520; AAA18942_1; -
 DR EMBL: U02520; AAA18943_1; -
 DR EMBL: AB008495; BAA25064_1; -
 DR MIM: 120070; -
 DR MIM: 203780; -
 DR MIM: 233450; -
 DR InterPro; IPR001442; C4
 DR Pfam; PF01413; C4; 2
 DR Pfam; PF01391; Collagen; 21
 DR SMART; SM00111; C4; 2
 DR ProDom; PD03933; C4; 2
 DR Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 KW Glycoprotein; Basement membrane; Collagen; Signal; Cell adhesion;
 KW Alternative splicing; Polymorphism; Phosphorylation; Disease mutation;
 KW Airport syndrome.
 FT SIGNAL 1 28 POTENTIAL.
 FT CHAIN 29 1670 COLLAGEN ALPHA 3(IV) CHAIN.
 FT DOMAIN 29 42 7S DOMAIN.
 FT DOMAIN 43 1438 TRIPLE-HELICAL REGION.
 FT DOMAIN 1439 1670 NONHELICAL REGION (NC1) (GOODPASTURE
 FT ANTIGEN) (BY SIMILARITY).
 FT DOMAIN 1427 1444 EPIPOLE (RECOGNIZED BY GOODPASTURE
 FT ANTIBODIES).
 FT SITE 1426 1427 CLEavage (BY COLLAGENASE)
 FT (BY SIMILARITY).
 SITE 791 793 CELL ATTACHMENT SITE (POTENTIAL).
 SITE 996 998 CELL ATTACHMENT SITE (POTENTIAL).

FT SITE 1154 1156 CELL ATTACHMENT SITE (POTENTIAL).
 FT SITE 1306 1308 CELL ATTACHMENT SITE (POTENTIAL).
 FT SITE 1345 1347 CELL ATTACHMENT SITE (POTENTIAL).
 FT SITE 1432 1434 CELL ATTACHMENT SITE (POTENTIAL).
 FT CARBOHYD 253 253 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT MOD_RES 1435 1435 PHOSPHORYLATION (BY SIMILARITY).
 FT DISULFID 1460 1551 PHOSPHORYLATION (BY SIMILARITY).
 FT DISULFID 1493 1548 OR 1548 (BY SIMILARITY).
 FT DISULFID 1505 1511 OR 1511 (BY SIMILARITY).
 FT DISULFID 1570 1665 BY SIMILARITY.
 FT DISULFID 1604 1662 OR 1662 (BY SIMILARITY).
 FT DISULFID 1616 1622 OR 1665 (BY SIMILARITY).
 FT VARSPLIC 1586 1670 BY SIMILARITY.
 FT VARSPLIC 1586 1670 BY SIMILARITY.
 FT VARSPLIC 1488 1670 BY SIMILARITY.
 FT VARSPLIC 1488 1670 BY SIMILARITY.
 FT VARIANT 43 43 /FTid=VAR_011202.
 FT VARIANT 162 162 G -> E.
 FT COAS_BPPRD STANDARD; PRT: 339 AA.
 AC P22536; ID CA1F_HUMAN
 DT 01-AUG-1991 (Rel. 19, Created)
 DT 01-DEC-1992 (Rel. 24, Last sequence update)
 DE Minor capsid protein (Protein P5).
 GN 1189 GAKGDRGAPGFFGGLP 1203
 RESULT 9
 COAS_BPPRD STANDARD; PRT: 339 AA.
 AC P22536; ID CA1F_HUMAN
 DT 01-AUG-1991 (Rel. 19, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Collagen alpha 1(XVII) chain precursor.
 GN 1189 GAKGDRGAPGFFGGLP 1203
 DB 123 GIKGDKGDPGAPG 135
 RESULT 10
 COAS_BPPRD STANDARD; PRT: 1603 AA.
 AC Q07092; ID CA1F_HUMAN
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Collagen alpha 1(XVII) chain precursor.
 GN 123 GIKGDKGDPGAPG 135
 OC Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TAXID=9606;
 RP [1] SEQUENCE FROM N.A.
 RX MEDLINE-92333339; PubMed=1631157;
 RA Pan T.-C., Zheng R. Z., Mattei M.-G., Timpl R., Chu M.-L.;
 RT "Cloning and chromosomal location of human alpha 1(XVI) collagen.";
 RL PROC. NATL. ACADEM. SCI. U.S.A. 89:6565-6569(1992).
 RN [2]
 RP SEQUENCE OF 418-1603 FROM N.A.
 RC TISSUE=Placenta;
 RX MEDLINE=93203161; PubMed=1284248;
 RA Yamaguchi N., Kimura S., McBride O.W., Hori H., Yamada Y.,
 RA Kanamori T., Yamakoshi H., Nagai Y.;
 RT "Molecular cloning and partial characterization of a novel collagen
 chain, alpha 1(XVI), consisting of repetitive collagenous domains and
 cysteine-containing non-collagenous segments.";
 RL J. Biochem. 112:856-861(1992).
 CC FUNCTION: THE NUMEROUS INTERRUPTIONS IN THE TRIPLE HELIX MAY MAKE
 CC THIS MOLECULE EITHER ELASTIC OR FLEXIBLE.
 CC -1- TISSUE SPECIFICITY: IN THE PLACENTA, WHERE IT IS FOUND IN THE
 CC AMNIOTIC FLUID, IT IS FOUND IN AN ACCELLULAR, RELATIVELY DENSE LAYER OF A
 CC COMPLEX NETWORK OF RETICULAR FIBERS, ALSO LOCATED TO A FIBROBLAST
 CC LAYER BEHIND THIS DENSE LAYER. EXISTS IN TISSUES IN ASSOCIATION
 CC WITH OTHER TYPES OF COLLAGEN.
 CC -1- DEVELOPMENTAL STAGE: TRANSIENTLY ELEVATED EXPRESSION DURING
 CC GESTATION AND DECREASE AT TERM.
 CC -1- DOMAIN: THIS SEQUENCE DEFINES EIGHTEEN DIFFERENT DOMAINS, NINE
 CC TRIPLE-HELICAL DOMAINS (COL9 TO COL1) AND TEN NONTRIPLE-HELICAL
 CC DOMAINS (NC10 TO NC1).
 CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPETIDE REPEATING
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 CC -1- SIMILARITY: BELONGS TO THE FIBRIL-ASSOCIATED COLLAGENS WITH
 CC INTERRUPTED HELICES (FACIT) FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBI; MG9077; AAA32460.1; --.
 DR EMBI; M55568; AAA32446.1; --.
 DR B46345; B46345.
 PIR;
 Coat protein.
 KW
 FT INIT_MET 0 0
 FT DOMAIN 1 121 DOMAIN-1
 FT DOMAIN 122 140 COLLAGEN-LIKE.
 FT DOMAIN 141 339 DOMAIN-2.
 SQ 339 AA: 34318 MW; 184BCE77D08944F3 CRC64;
 Query Match 66 3%; Score 61; DB 1; Length 339;
 Best Local Similarity 75 9%; Pred. No. 0.046;
 Matches 10; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -	OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
CC	the European Bioinformatics Institute. There are no restrictions on its	OC	Mammalia; Eutheria; Primates; Catarhini; Hominidae; Homo.	
CC	use by non-profit institutions as long as its content is in no way	OC		
CC	modified and this statement is not removed. Usage by and for commercial	NCBI_TaxID=9606;		
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/	[1]		
CC	or send an email to license@isb-sib.ch).	SEQUENCE FROM N.A.	TISSUE-Kidney.	
CC		RC	RC	
DR	InterPro: IPR003129; TSPN:	RX	MEDLINE=9501445; PubMed=7523402;	
DR	pfam; PF01391; Collagen; 1.	RX	Leinonen A., Mariyama M., Mochizuki T., Tryggvason K., Reenders S.T.,	
DR	SMART; SM00210; TSPN: 1.	RX	Complete primary structure of the human type IV collagen alpha 4(IV) chain. Comparison with structure and expression of the other alpha	
DR		RX	(IV) chains.";	
KW	Extracellular matrix; Connective tissue; Collagen; Hydroxylation;	RN	J. Biol. Chem. 269:26172-26177(1994).	
KW	Repair; Signal.	RN	SEQUENCE OF 1-223 FROM N.A.	
FT	SIGNAL 1 21	RN	[2]	
FT	CHAIN 22 1603	RN	SEQUENCE OF 1-223 FROM N.A.	
FT	DOMAIN 22 374	RN	SEQUENCE OF 1-223 FROM N.A.	
FT	DOMAIN 375 505	RN	SEQUENCE OF 1-223 FROM N.A.	
FT	DOMAIN 506 520	RN	SEQUENCE OF 1219-1630 FROM N.A.	
FT	DOMAIN 521 554	RN	SEQUENCE OF 1219-1630 FROM N.A.	
FT	DOMAIN 555 571	RN	SEQUENCE OF 1219-1630 FROM N.A.	
FT	DOMAIN 572 630	RN	SEQUENCE OF 1219-1630 FROM N.A.	
FT	DOMAIN 631 651	RN	SEQUENCE OF 1219-1630 FROM N.A.	
FT	DOMAIN 652 722	RN	SEQUENCE OF 1407-1507 FROM N.A.	
FT	DOMAIN 723 737	RN	SEQUENCE OF 1407-1507 FROM N.A.	
FT	DOMAIN 738 875	RN	SEQUENCE OF 1407-1507 FROM N.A.	
FT	DOMAIN 876 896	RN	SEQUENCE OF 1407-1507 FROM N.A.	
FT	DOMAIN 887 938	RN	SEQUENCE OF 1407-1507 FROM N.A.	
FT	DOMAIN 939 972	RN	SEQUENCE OF 1407-1507 FROM N.A.	
FT	DOMAIN 973 987	RN	SEQUENCE OF 1407-1507 FROM N.A.	
FT	DOMAIN 988 1010	RN	SEQUENCE OF 1407-1507 FROM N.A.	
FT	DOMAIN 1011 1432	RN	SEQUENCE OF 1407-1507 FROM N.A.	
FT	DOMAIN 1433 1471	RN	SEQUENCE OF 1407-1507 FROM N.A.	
FT	DOMAIN 1472 1577	RN	SEQUENCE OF 1407-1507 FROM N.A.	
FT	DOMAIN 1578 1603	RN	SEQUENCE OF 1407-1507 FROM N.A.	
FT	CONFLICT 418 420	RN	SEQUENCE OF 1407-1507 FROM N.A.	
FT	CONFLICT 537 537	RN	SEQUENCE OF 1407-1507 FROM N.A.	
FT	CONFLICT 1160 1160	RN	SEQUENCE OF 1407-1507 FROM N.A.	
FT	CONFLICT 1163 1163	RN	SEQUENCE OF 1407-1507 FROM N.A.	
FT	CONFLICT 1165 1165	RN	SEQUENCE OF 1407-1507 FROM N.A.	
SQ	SEQUENCE 1603 AA; E27D9A1D4E598A37 CRC64;	RN	VARIANT BEH GJU-897.	
OS	Query Match 66.3%; Score 61; DB 1; Length 1603;	RX	MEDLINE=9637960; PubMed=8787673;	
OS	Best Local Similarity 73.3%; Pred. No. 0.21;	RX	Leemink H.H., Millesen W.N., Mochizuki T., Schroeder C.H.,	
OS	Matches 11; Conservative 0; Mismatches 4; Indels 0; Gaps 0;	RX	Brunner H.G., van Oost B.A., Monnens L.A.H., Smeets H.J.M.;"	
QY	1 GVKGDKGNPGWPAG 15	RX	Benign familial hematuria due to mutation of the type IV collagen	
Db	681 QOKGDAGNPDPGTP 695	RX	alpha gene.";	
OS	Homo sapiens (Human).	RN	J. Clin. Invest. 98:1114-1118(1995).	
RESULT 11		RN	[8]	
CA4-HUMAN	STANDARD; PRT; 1690 AA.	RX	VARIANTS AS, AND VARIANTS.	
ID		RX	MEDLINE=99011253; PubMed=9792860;	
AC		RA	Boye E., Mollet G., Forestier L., Cohen-Solal L., Heide L.,	
AC	P53420; 01-Oct-1996 (Rel. 34. Created)	RA	Cochet P., Grunfeld J.-P., Palcoux J.-B., Gubler M.-C., Antignac C.;	
DT	01-Oct-1995 (Rel. 34. Last sequence update)	RA	"Determination of the genomic structure of the COL4A4 gene and of	
DT	01-MAR-2002 (Rel. 41. Last annotation update)	RT	novel mutations causing autosomal recessive Alport syndrome.";	
DE	Collagen alpha 4(IV) chain precursor.	RT	Am. J. Hum. Genet. 63:1219-1240(1998).	
GN	COL4A4.	RT	-!- FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF	
OS		CC	GLOMERULAR BASEMENT MEMBRANES (GBM), FORMING A 'CHICKEN WIRE'	
		CC	MESHWORK TOGETHER WITH LAMININS, PROTEOGLYCANS AND ENACTIN/	
		CC	NIDogen.	
		CC	-!- SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOFORMS, ALPHA 1(IV)-	

CC ALPHA 6(IV), EACH OF WHICH CAN FORM A TRIPLE HELIX STRUCTURE WITH 2 OTHER CHAINS TO GENERATE TYPE IV COLLAGEN NETWORK.

CC -1- SUBCELLULAR LOCATION: CELL SURFACE (POTENTIAL).

CC -1- TISSUE SPECIFICITY: ALPHA 3 AND ALPHA 4 TYPE IV COLLAGENS ARE COLOCALIZED AND PRESENT ONLY IN BASEMENT MEMBRANES OF KIDNEY, EYE, COCHLEA, LUNG AND BRAIN.

CC -1- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL TRIPLE-HELICAL 7S DOMAIN.

CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPETIDE REPEATING UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.

CC -1- PTM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH ARE INVOLVED IN INNER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF THESE, LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE IV COLLAGENS.

CC -1- DISEASE: DEFECTS IN COL4A4 ARE A CAUSE OF FAMILIAL BENIGN RECESSIVE FORM OF ALPORT SYNDROME (AS), AN HEREDITARY GLOMERULONEPHROPATHY CHARACTERIZED BY PROGRESSIVE RENAL FAILURE, HEMATURIA AND DEAFNESS. THE RECESSIVE FORM OCCURS EQUALLY BETWEEN MALES AND FEMALES.

CC -1- DISEASE: DEFECTS IN COL4A4 ARE A CAUSE OF FAMILIAL BENIGN HEMATURIA (FBB) OR THIN BASEMENT MEMBRANE DISEASE. FBB IS CHARACTERIZED BY PERSISTENT HEMATURIA, AN ELECTRON MICROSCOPICALLY DETECTABLE THIN GLOMERULAR BASEMENT MEMBRANE (GBM) AND AN AUTOSOMAL DOMINANT MODE OF INHERITANCE. RENAL FUNCTION REMAINS NORMAL. IN CHILDREN DIFFERENTIATION BETWEEN FBB AND AS CAN BE DIFFICULT, BECAUSE BOTH DISORDERS ARE MANIFESTED BY PERSISTENT HEMATURIA AND THIN GBM AT THAT AGE.

CC -1- SIMILARITY: TO OTHER TYPE IV COLLAGENS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement. (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC

CC EMBL; X80103; CA56943.1; -. DR EMBL; AB008496; BA25065.1; -. DR EMBL; D17391; BAA04214.1; -. DR EMBL; AB013131; -. DR MIM; 141200; -. DR MIM; 203780; -. DR InterPro; IPR001442; C4. DR InterPro; IPR000087; C4. DR Pfam; PF01413; C4; 2. DR Pfam; PF01391; Collagen; 20. DR SMART; SM0011; C4; 2. DR SWISS-PROT; SW0011; C4; 2. KW Extracellular matrix; Connective tissue; Basement membrane; Repeat; Polymorphism; Alport syndrome.

CC SIGNAL 1 38 POTENTIAL.

FT DOMAIN 39 1690 COLLAGEN ALPHA 4(IV) CHAIN.

FT DOMAIN 39 64 7S DOMAIN.

FT DOMAIN 65 1459 TRIPLE-HELICAL REGION.

FT DOMAIN 1460 1690 NONHELICAL REGION (NC1).

FT SITE 94 96 CELL ATTACHMENT SITE (POTENTIAL).

FT SITE 145 147 CELL ATTACHMENT SITE (POTENTIAL).

FT SITE 189 191 CELL ATTACHMENT SITE (POTENTIAL).

FT SITE 310 312 CELL ATTACHMENT SITE (POTENTIAL).

FT SITE 724 726 CELL ATTACHMENT SITE (POTENTIAL).

FT SITE 785 787 CELL ATTACHMENT SITE (POTENTIAL).

FT SITE 989 991 CELL ATTACHMENT SITE (POTENTIAL).

FT SITE 1206 1206 COLLAGENASE (BY SIMILARITY).

FT SITE 1212 1214 CELL ATTACHMENT SITE (POTENTIAL).

FT DISULFID 1480 1569 DISULFIDE (BY SIMILARITY).

FT DISULFID 1513 1566 DISULFIDE (BY SIMILARITY).

FT DISULFID 1525 1531 BY SIMILARITY.

FT DISULFID 1588 1686 OR 1683 (BY SIMILARITY).

FT DISULFID 1622 1683 OR 1686 (BY SIMILARITY).

FT DISULFID 1634 1641 BY SIMILARITY.

FT CARBOHYD 142 142 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 669 669 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT VARIANT 441 446 MISSING (IN AS).

FT VARIANT 446 /FRId=VAR_008148.

FT VARIANT 545 545 G -> A.

FT VARIANT 570 570 /FRId=VAR_008149.

FT VARIANT 570 570 E -> Q.

FT VARIANT 570 570 L -> P.

FT VARIANT 570 570 /FRId=VAR_008150.

FT VARIANT 897 897 G -> E (IN FBB).

FT VARIANT 931 931 /FRId=VAR_001912.

FT VARIANT 1201 1201 A -> T.

FT VARIANT 1402 1402 /FRId=VAR_008151.

FT VARIANT 1402 1402 P -> S.

FT VARIANT 1572 1572 /FRId=VAR_008154.

FT VARIANT 1572 1572 P -> L (IN AS).

FT CONFLICT 1659 1660 /FRId=VAR_008155.

FT SEQUENCE 1690 AA; 164095 MW; LQ -> FE (IN REF. 3).

FT SEQUENCE 1690 AA; 164095 MW; E1E2F2B3A72BAAE CRC64;

FT SEQUENCE 184 GIQGDGDGPGLP 198

Query Match 66.3%; Score 61; DB 1; Length 1690;

Best Local Similarity 60.0%; Pred. No. 0; 22; Mismatches 4;保守性 9; Conservative 4; Matches 9; Indels 2; Gaps 0; Gaps 0;

Db

RESULT 12

ID	CA24_ASCSU	STANDARD	PRT:	1763 AA.
AC	C24_ASCSU			
DT	P21393; 01-AUG-1992 (Rel. 23, Created)			
DT	01-AUG-1992 (Rel. 23, Last sequence update)			
DT	30-MAY-2000 (Rel. 39, Last annotation update)			
DE	Collagen alpha 2(IV) chain precursor.			
OS	Ascaris suum (Pig roundworm) (Ascaris lumbricoides)			
OC	Eukaryota; Metazoa; Nemata; Chromadorea; Ascarididae; Ascaridoidea; Ascarididae; Ascariidae.			
OX	NCBI_TaxID=6233;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	MEDLINE-91340768; PubMed=1714907;			
RT	"The complete primary structure of a nematode alpha 2(IV) collagen and the partial structural organization of its gene.";			
RL	J. Biol. Chem. 26:1649-1615(1991).			
CC	-1- FUNCTION: COLLAGEN TYPE IV IS SPECIFIC FOR BASEMENT MEMBRANES.			
CC	-1- SUBUNIT: TIMERS OF TWO ALPHA 1(IV) AND ONE ALPHA 2(IV) CHAIN.			
CC	-1- TYPE IV COLLAGEN FORMS A MESH-LIKE NETWORK LINKED THROUGH INTERMOLECULAR INTERACTIONS BETWEEN 7S DOMAINS AND BETWEEN NC1 DOMAINS.			
CC	-1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: I (SHOWN HERE) AND II; ARE PRODUCED BY ALTERNATIVE SPLICING.			
CC	-1- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL TRIPLE-HELICAL 7S DOMAIN.			
CC	-1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPETIDE REPEATING UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.			
CC	-1- PTM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH ARE INVOLVED IN INNER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF THESE, LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE IV COLLAGENS.			

CC IV COLLAGENS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC DR EMBL; M67507; AAA1014.1; .

DR PIR; S16366; S16366.

DR InterPro; IPR001442; C4.

DR InterPro; IPR00087; Collagen.

DR Pfam; PF01413; C4; 2.

DR Pfam; PF01391; Collagen; 23.

DR Prodom; PD003923; C4; 2.

DR SMART; SM00111; C4; 2.

KW Hydroxylation; Connective tissue; Basement membrane; Repeat; Collagen;

KW Alternative splicing; Glycoprotein; Signal.

KW SIGNAL 1

FT CHAIN 27 1763

FT DOMAIN 27 42

FT DOMAIN 43 1529

FT DOMAIN 1530 1763

FT DISUFLID 1548 1637

FT DISUFLID 1581 1634

FT DISUFLID 1593 1599

FT DISUFLID 1555 1752

FT DISUFLID 1690 1749

FT DISUFLID 1702 1709

FT CARBOHYD 125 126

FT CARBOHYD 249 249

VARSPLIC 230 266

FT SQ SEQUENCE 1763 AA; 168526 MW; 304F538BC056AAE0D CRC64;

Query Match 66.3%; Score 61; DB 1; Length 1763;

Best Local Similarity 73.3%; Pred. No. 0.23;

Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 GVKGDKGNGNPGAP 15

Db 760 I:||||| 151 GLPGMKGNPGLPGAP 774

RESULT 13

PSPD_BOVIN ID PSPD_BOVIN STANDARD; PRT; 369 AA.

AC P35246;

DT 01-FEB-1994 (Rel. 28, Created)

DT 01-FEB-1994 (Rel. 28, last sequence update)

DT 15-JUN-1999 (Rel. 38, last annotation update)

DE Pulmonary surfactant-associated protein D precursor (SP-D) (PSP-D).

GN SFTP4.

OS Bos taurus (Bovine).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bovinae; Bos.

NCBI_TAXID=9913;

[1] SEQUENCE FROM N.A., AND SEQUENCE OF 208-247.

TISSUE=Lung;

MEDLINE=33170856; PubMed=8434402;

RA Lim B.L.; Lu J.; Reid K.B.M.;

RA "Structural similarity between bovine conglutinin and bovine lung surfactant protein D and demonstration of liver as a site of synthesis of conglutinin.";

RT Immunology 78:159-165 (1993).

CC -I- FUNCTION: CONTRIBUTES TO THE LUNG'S DEFENSE AGAINST INHALED MICROORGANISMS. BINDS STRONGLY MALTOSE RESIDUES AND TO A LESSER

CC EXTENT OTHER ALPHA-GLUCOSYL MOieties. IT COULD PARTICIPATE IN THE EXTRACELLULAR REORGANIZATION OR TURNOVER OF PULMONARY SURFACTANT.

CC -I- SUBUNIT: OLIGOMERIC COMPLEX OF 4 SET OF HOMOTRIMERS.

CC -I- SUBCELLULAR LOCATION: Extracellular.

CC -I- MISCELLANEOUS: PULMONARY SURFACTANT CONSISTS OF 90% LIPID AND 10% PROTEIN. THERE ARE 4 SURFACTANT ASSOCIATED PROTEIN: 2 COLLAGENOUS, HYDROPHOBIC PROTEINS (SP-B AND SP-C).

CC -I- SIMILARITY: CONTAINS 1 COLLAGENOUS DOMAIN.

CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC DR EMBL; X7911; CAA3510.1; .

CC DR PIR; S33603; S33603.

CC DR HSSP_P35247; 1B08.

CC DR InterPro; IPR00087; Collagen.

CC DR InterPro; IPR001304; Lectin_c.

CC DR Pfam; PF01391; Collagen; 3.

CC DR Pfam; PF00059; Lectin_c; 1.

CC DR SMART; SM0034; CLCT; 1.

CC DR PROSTE; PS00615; C-TYPE LECTIN_1; 1.

CC DR PROSTE; PS50041; C-TYPE LECTIN_2; 1.

CC KW Glycoprotein; Calcium; Surface film; Gaseous exchange; Hydroxylation; Signal; Lectin; Collagen; Repeat; Coiled coil.

CC FT DOMAIN 217 248

CC FT DOMAIN 273 369

CC FT DISUFLID 345 359

CC FT CARBOHYD 90 90

CC FT MOD_RES 78 78

CC FT MOD_RES 87 87

CC FT MOD_RES 96 99

CC FT MOD_RES 99 99

CC FT MOD_RES 165 165

CC FT MOD_RES 171 171

CC SQ SEQUENCE 369 AA; 37361 MW; 07D8B24E0ABE23 CRC64;

Query Match 65.2%; Score 60; DB 1; Length 369;

Best Local Similarity 66.7%; Pred. 0.07;

Matches 10; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GVKGDKGNGNPGAP 15

Db 157 GLKGERGAPEEPGAP 171

RESULT 14

CA44_RABIT ID CA44_RABIT STANDARD; PRT; 623 AA.

AC P55787;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, last sequence update)

DT 30-MAY-2000 (Rel. 39, last annotation update)

DE Collagen alpha 4(IV) chain (Fragment).

GN COL4A4.

OS Oryctolagus cuniculus (Rabbit).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.

OX NCBL_TAXID=9986;

RN [1] SEQUENCE FROM N.A.

RC TISSUE=Corneal endothelium;

RX MEDLINE=93054733; PubMed=1429714;
 RA Kamagata Y., Mattel M.-G., Niromiya Y.;
 RT "Isolation and sequencing of cDNAs encoding the
 alpha 4 chain of basement membrane collagen type IV and assignment of
 the gene to the distal long arm of human chromosome 2.";
 RL J. Biol. Chem. 267:23753-23758(1992).
 CC -I- FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF
 GLOBULAR BASEMENT MEMBRANES (GBM), FORMING A 'CHICKEN-WIRE'
 MESHWORK TOGETHER WITH LAMININS, PROTEOGLYCANS AND ENTHACTIN/
 NIDGEN. NIDGEN.
 CC -I- SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOFORMS, ALPHA 1(IV)-
 ALPHA 6(IV), EACH OF WHICH CAN FORM A TRIPLE HELIX STRUCTURE
 WITH 2 OTHER CHAINS TO GENERATE TYPE IV COLLAGEN NETWORK.
 CC -I- SUBCELLULAR LOCATION: CELL SURFACE (POTENTIAL).
 CC -I- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS
 DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE
 G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY
 CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL
 TRIPLE-HELICAL 7S DOMAIN.
 CC -I- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPETIDE REPEATING
 UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 CC -I- PTM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH
 ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF
 THESE LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE
 IV COLLAGENS.
 CC -I- SIMILARITY: TO OTHER TYPE IV COLLAGENS.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC DR EMBL; L01477; ; NOT_ANNOTATED_CDS.
 CC DR InterPro; IPR00142; C4
 CC DR InterPro; IPR00007; Collagen.
 CC DR Pfam; PF01413; C4; 2.
 CC DR ProDom; PD003923; C4; 2.
 CC DR SMART; SM00111; C4; 2.
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 KW Glycoprotein; Basement membrane; Collagen; Cell adhesion.
 FT NON_TER 1 1 TRIPLE-HELICAL REGION.
 FT DOMAIN <1 392 NONHELICAL REGION (NC1).
 FT DISULFID 393 623 OR 499 (BY SIMILARITY).
 FT DISULFID 413 502 OR 502 (BY SIMILARITY).
 FT DISULFID 446 499 BY SIMILARITY.
 FT DISULFID 458 464 BY SIMILARITY.
 FT DISULFID 521 619 OR 616 (BY SIMILARITY).
 FT DISULFID 555 616 BY SIMILARITY.
 FT DISULFID 567 574 BY SIMILARITY.
 SQ SEQUENCE 623 AA; MW: 62393 MW: CCGC9B931242FE82 CRC64:
 Query Match 65.2%; Score 60; DB 1; Length 623;
 Best Local Similarity 76.9%; Pred. No. 0.12; 1; Mismatches 10; Conservative 1; Indels 0; Gaps 0;
 Matches 10; Non-termini 1; Carboxyl-terminal 1; Conflicts 0;

Oy 1 GVKGDKGNPGWPG 13
 1:||||||| ||||| 139 GIKGDKGKGPSPG 151

RESULT 15
 CA13_RAT STANDARD: PRT; 636 AA.
 AC P13941; 070604;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Collagen alpha 1(III) chain (Fragment).
 GN COL3A1.

OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mamalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TAXID=10116;
 RN [1] SEQUENCE FROM N.A.
 RX MEDLINE=94114571; PubMed=8286415;
 RA Glunhoff V., Maekelae J. K., Vuorio E.;
 RT Cloning of cDNA for rat pro alpha 1(III) collagen mRNA. Different
 expression patterns of type I and type III collagen and fibronectin
 genes in experimental granulation tissue.;
 RT Biocell. Biophys. Acta 127:41-48(1994).
 RN [2] SEQUENCE OF 73-636 FROM N.A.
 RX STRAINS-Sprague-Dawley; TISSUE=FBroblast;
 RA Wurtz T., Ellerstroem C., Lundmark C., Christersson C.;
 RA Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
 RN [3] SEQUENCE OF 308-482 FROM N.A.
 RX MEDLINE=88296083; PubMed=2456904;
 RA Franken F.R., Hsu C.-Y.J., Meyers J.C., Lin E., Lytle C.R.,
 RA Komn B., Mohn K.;
 RA Regulation of alpha 2(II), alpha 1(III), and alpha 2(V) collagen
 RNA by estradiol in the immature rat uterus.;
 RL DNA 7:347-356(1988).
 CC -I- FUNCTION: COLLAGEN TYPE III OCCURS IN MOST SOFT CONNECTIVE TISSUES
 CC ALONG WITH TYPE I COLLAGEN.
 CC -I- SUBUNIT: TRIMERS OF IDENTICAL ALPHA 1(III) CHAINS. THE CHAINS ARE
 CC LINKED TO EACH OTHER BY INTERCHAIN DISULFIDE BONDS. TRIMERS ARE
 CC ALSO CROSS-LINKED VIA HYDROXYLYSINES.
 CC -I- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPETIDE REPEATING
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 CC
 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC DR EMBL; X70369; CAM49832; 1; -.
 CC DR EMBL; AJ005395; CAA06510; 1; -.
 CC DR EMBL; M21354; AAC00942; 1; -.
 CC PIR; A29905; A29905.
 CC PIR; S41067; S41067.
 CC DR InterPro; IPR000087; Collagen.
 CC DR InterPro; IPR00805; Fib_collagen_C.
 CC DR InterPro; IPR01007; WFEC.
 CC DR Pfam; PF01410; COLFI; 1.
 CC DR ProDom; PD002078; Fib_collagen_C; 1.
 CC DR SMART; SM00038; COLFI; 1.
 CC DR PROSITE; PS01208; WFEC; PARTIAL.
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 KW Collagen; Glycoprotein.
 FT NON_TER 1 1 COLLAGEN ALPHA 1(III) CHAIN.
 FT CHAIN <1 375 CARBOXYL-TERMINAL PROPEPTIDE.
 FT PROPEP 376 636 TRIPLE-BILICAL REGION.
 FT DOMAIN <1 368 NONHELICAL REGION (C-TERMINAL).
 FT DISULFID 369 636 INTERCHAIN (BY SIMILARITY).
 FT DISULFID 369 369 INTERCHAIN (BY SIMILARITY).
 FT CONFLICT 340 340 N -> D (IN REF. 2).
 FT CONFLICT 429 429 A -> G (IN REF. 2).
 SQ SEQUENCE 636 AA; MW: 6148159F01D001EE CRC64:
 Query Match 65.2%; Score 60; DB 1; Length 636;
 Best Local Similarity 71.4%; Pred. No. 0.12; 1; Mismatches 3; Conservative 3; Indels 0; Gaps 0;
 Matches 10; Non-termini 1; Carboxyl-terminal 1; Conflicts 0;

Oy 1 GVKGDKGNPGWPG 14
 ||||:||||||| 139 GIKGDKGKGPSPG 151

Db: 30 GVKGGERGSPGGPA 43

RESULT 16
 CAL13_BOVIN ID CAL13_BOVIN STANDARD; PRT; 1049 AA.
 AC P04238;
 DT 20-MAR-1997 (Rel. 04, Created)
 DT 20-MAR-1997 (Rel. 04, Last sequence update)
 DE 16-OCT-2001 (Rel. 40, Last annotation update)
 DE collagen alpha 1(III) chain.
 GN COL3A1.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea; Bovidae; Bovinae; Bos.
 NCBI_TAXID=9913;
 [1]
 RP SEQUENCE OF 1-242; PubMed=488006;
 RX MEDLINE=80026026;
 RA Fietzek P.P., Alimann H., Rauterberg J., Henkel W., Wachter E.,
 RA Kuhn K.;
 RT "The covalent structure of calf skin type III collagen. I. The amino acid sequence of the amino terminal region of the alpha 1(III) chain (positions 1-222).";
 RT Hoppe-Seyler's Z. Physiol. Chem. 360:809-820(1979).
 RL [2]
 RP SEQUENCE OF 243-422;
 RX MEDLINE=80026027; PubMed=488007;
 RA Dewes H., Fietzek P.P., Kuhn K.;
 RT "The covalent structure of calf skin type III collagen. II. The amino acid sequence of the cyanogen bromide peptide alpha 1(III)CB1,8,10,2 (positions 223-402).";
 RT Hoppe-Seyler's Z. Physiol. Chem. 360:821-832(1979).
 RL [3]
 RP SEQUENCE OF 423-571.
 RX MEDLINE=80026028; PubMed=488008;
 RA Bentz H., Fietzek P.P., Kuhn K.;
 RT "The covalent structure of calf skin type III collagen. III. The amino acid sequence of the cyanogen bromide peptide alpha 1(III)CB4 (positions 403-551).";
 RT Hoppe-Seyler's Z. Physiol. Chem. 360:833-840(1979).
 RL [4]
 RP SEQUENCE OF 572-808.
 RX MEDLINE=80026029; PubMed=488009;
 RA Lang H., Glanville R.W., Fietzek P.P., Kuhn K.;
 RT "The covalent structure of calf skin type III collagen. IV. The amino acid sequence of the cyanogen bromide peptide alpha 1(III)CB5 (positions 552-788).";
 RT Hoppe-Seyler's Z. Physiol. Chem. 360:841-850(1979).
 RL [5]
 RP SEQUENCE OF 809-947.
 RX MEDLINE=80026030; PubMed=488010;
 RA Dewes H., Fietzek P.P., Kuhn K.;
 RT "The covalent structure of calf skin type III collagen. V. The amino acid sequence of the cyanogen bromide peptide alpha 1(III)CB9A (position 789-927).";
 RT Hoppe-Seyler's Z. Physiol. Chem. 360:851-860(1979).
 RL [6]
 RP SEQUENCE OF 948-1049.
 RX MEDLINE=80026031; PubMed=488011;
 RA Allmann H., Fietzek P.P., Glanville R.W., Kuhn K.;
 RT "The covalent structure of calf skin type III collagen. VI. The amino acid sequence of the carboxy-terminal cyanogen bromide peptide alpha 1(III)CB9B (positions 928-1028).";
 RT Hoppe-Seyler's Z. Physiol. Chem. 360:861-868(1979).
 RL [7]
 RP SEQUENCE: COLLAGEN TYPE III OCCURS IN MOST SOFT CONNECTIVE TISSUES ALONG WITH TYPE I COLLAGEN
 CC -1- SUBUNIT: TRIMERS OF IDENTICAL ALPHA 1(III) CHAINS. THE CHAINS ARE LINKED TO EACH OTHER BY INTERCHAIN DISULFIDE BONDS. TRIMERS ARE ALSO CROSS-LINKED VIA HYDROXYLINES.
 CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPÉPTIDE REPEATING UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.

DR PIR: A02862; CCBO75;
 DR InterPro; IPR000087; Collagen.
 DR IntePro; IPR001007; WNC.
 DR Pfam; PF01391; Collagen; 17.
 DR PROSITE; PS01208; VWFC; PARTIAL.
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation; GLYCOPROTEIN; Collagen.
 FT DOMAIN 1 14 NONHELCICAL REGION (N-TERMINAL).
 FT DOMAIN 15 1040 TRIPLE-HELICAL REGION.
 FT DOMAIN 1041 1049 NONHELCICAL REGION (C-TERMINAL).
 FT MOD_RES 95 95 HYDROXYLATION.
 FT MOD_RES 107 107 HYDROXYLATION.
 FT CARBOHYD 950 950 O-LINKED (GAL. . .).
 FT CARBOHYD 950 950 O-LINKED (GAL. . .).
 FT DISUFRID 1040 1040 INTERCHAIN.
 FT DISUFRID 1041 1041 INTERCHAIN.
 SQ SEQUENCE 1049 AA; 93651 MW; 8EBC33DIC66ECC9A3 CRC64;
 QY 1 GVKGGERGSPGGPA 14
 ||||:||||| 715
 Db 702 GVKGGERGSPGGPA 715

RESULT 17
 CAL13_CHICK ID CAL13_CHICK STANDARD; PRT; 1262 AA.
 AC P12105; P79758; P79759; 090794; 092029;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Collagen alpha 1(III) chain precursor (Fragments).
 GN COL3A1.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
 OC NCBI_TAXID=9031;
 RN [1]
 RP SEQUENCE OF 1-886 FROM N.A.
 RC TISSUE=Kidney;
 RX MEDLINE=94266842; PubMed=8206952;
 RA Nah H.-D., Niu Z., Adams S.L.;
 RT "An alternative transcript of the chick type III collagen gene that does not encode type III collagen.";
 RL J. Biol. Chem. 269:16443-16448(1994).
 RN [2]
 RP SEQUENCE OF 29-96; 332-397; 431-484; 503-535 AND 869-976 FROM N.A.
 RX MEDLINE=84270596; PubMed=6547770;
 RA Yamada Y., Liou G., Mudryj M., Obici S., de Crombrugge B.;
 RT "Conservation of the sizes for one but not another class of exons in two chick collagen genes.";
 RL Nature 310:333-337(1984).
 RN [3]
 RP SEQUENCE OF 977-1262 FROM N.A.
 RX MEDLINE=83220819; PubMed=6855474;
 RA Yamada Y., Kuhn K., de Crombrugge B.;
 RT "A conserved nucleotide sequence, coding for a segment of the C-propeptide, is found at the same location in different collagen genes.";
 RT Nucleic Acids Res. 11:2733-2744(1983).
 CC -1- FUNCTION: COLLAGEN TYPE III OCCURS IN MOST SOFT CONNECTIVE TISSUES
 CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPÉPTIDE REPEATING UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.

CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 CC -1- SIMILARITY: CONTAINS 1 WIFC DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration -
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC use by non-profit Institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC

DR EMBL; U07973; AAA83407; 1.
 DR EMBL; X00822; CAB52686; 1.
 DR EMBL; X00823; CAB52686; 1; JOINED.
 DR EMBL; X00826; CAA25397; 1; ALT_Seq.
 DR EMBL; X00825; CAA25397; 1; JOINED.
 DR EMBL; X00827; CAA25398; 1; -.
 DR EMBL; X00828; CAA25399; 1; -.
 DR EMBL; X00830; CAA25401; 1; -.
 DR EMBL; X00831; CAA25402; 1; -.
 DR EMBL; K02302; AAD15299; 1; -.
 DR PRODOM; PD002078; Fib.collagen_C; 1.
 DR PIR; A05269; A05269.
 DR InterPro; IPR00087; Collagen.
 DR InterPro; IPR00085; Fib.collagen_C.
 DR InterPro; IPR001007; WIFC.
 DR Pfam; PF0191; Collagen; 13.
 DR PRODOM; PD002078; Fib.collagen_C; 1.
 DR SMART; SM0038; COLFI; 1.
 DR SMART; SM00214; WIFC; 1.
 DR PROSITE; PS01208; WIFC; 1.
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 KW Glycoprotein; Collagen; Signal.
 FT SIGNAL 1 23 AMINO-TERMINAL PROPEPTIDE (BY POTENTIAL).
 FT PROPERP 24 144 AMINO-TERMINAL PROPEPTIDE (BY SIMILARITY).
 FT CHAIN 145 1003 COLLAGEN ALPHA 1(III) CHAIN.
 FT PROPEP 1004 1262 CARBOXYL-TERMINAL PROPEPTIDE (BY SIMILARITY).
 FT DOMAIN 29 88 VWF_C.
 FT DOMAIN 145 164 NONHELICAL REGION (N-TERMINAL) (BY SIMILARITY).
 FT DOMAIN 165 994 TRIPLE-HELICAL REGION (BY SIMILARITY).
 FT DOMAIN 995 1003 NONHELICAL REGION (C-TERMINAL) (BY SIMILARITY).
 FT NON_CONS 886 887 INTERCHAIN (BY SIMILARITY).
 FT NON_CONS 922 923 INTERCHAIN (BY SIMILARITY).
 FT DISULFID 994 994 INTERCHAIN (BY SIMILARITY).
 FT DISULFID 995 995 HYDROXYLATION (BY SIMILARITY).
 FT MOD_RES 283 283 HYDROXYLATION (BY SIMILARITY).
 FT MOD_RES 859 859 HYDROXYLATION (BY SIMILARITY).
 FT CARBOHYD 1163 N-LINKED (GLCNAC, . .) (POTENTIAL).
 FT CONFLICT 96 96 E -> K (IN REF. 2).
 FT CONFLICT 1132 1132 F -> S (IN REF. 3).
 SQ SEQUENCE 1262 AA; 121249 MW; 96ABE7B2E9DEB43D CRC64;

Query Match 65.2%; Score 60; DB 1; Length 1262;
 Best Local Similarity 71.4%; Pred. No. 0; 24; Matches 10; conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 GVGDKGNPGWPCA 14
 QY ||||:||||| 14
 DB 857 GVGERGSPGGCA 870

RESULT 18
 CA13_HUMAN STANDARD; PRM; 1466 AA.
 AC P02461; Q15112; (Rel. 01, Created)
 DT 21-JUL-1996 (Rel. 01, Last sequence update)

DT 16-OCT-2001 (Rel. 40, last annotation update)
 DE Collagen alpha 1(III) chain precursor.
 GN COL3A1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrini; Hominoidea; Homo.
 OC NEBI_TAXID=9606;
 RN [1].
 RP SIMQUENCE FROM N.A.
 RC TISSUE-Skin fibroblast;
 RX MEDLINE=89310838; PubMed=2764886;
 RA Ala-Kotko L., Kontusari S., Baldwin C.T., Kulvantient H.,
 RA Prockop D.J.;
 RT Structure of cDNA clones coding for the entire prepro alpha 1 (III) chain of human type III procollagen. Differences in protein structure from type I procollagen and conservation of codon preferences.;
 RT Blochim. J. 260:509-516(1989).
 RL [2].
 RP SIMQUENCE OF 149-1225 FROM N.A.
 RX MEDLINE=89386015; PubMed=2780304;
 RA Janecko R.A., Ramirez F.;
 RT "Nucleotide and amino acid sequences of the entire human alpha 1 (III) collagen.";
 RT Nucleic Acids Res. 17:6742-6742(1989).
 RN [3].
 RP SIMQUENCE OF 168-398.
 RX MEDLINE=77134724; PubMed=557335;
 RA Seyer J.M., Kang A.H.;
 RT "Covalent structure of collagen: amino acid sequence of cyanogen bromide peptides from the amino-terminal segment of type III collagen of human liver.";
 RT Biochemistry 16:1158-1164(1977).
 RN [4].
 RP REVISIONS.
 RA Seyer J.M.;
 RA Submitted (DEC-1977) to the PIR data bank.
 RL [5].
 RP SIMQUENCE OF 399-727.
 RX MEDLINE=79000343; PubMed=687591;
 RA Seyer J.M., Kang A.H.;
 RT "Covalent structure of collagen: amino acid sequence of five consecutive CNBr peptides from type III collagen of human liver.";
 RT Biochemistry 17:3404-3411(1978).
 RL [6].
 RP SIMQUENCE OF 728-964.
 RX MEDLINE=80198282; PubMed=6246925;
 RA Seyer J.M., Mainardi C., Kang A.H.;
 RT "Covalent structure of collagen: amino acid sequence of alpha 1 (III)-CB5 from type III collagen of human liver.";
 RL Biochemistry 19:1583-1589(1980).
 RN [7].
 RP SIMQUENCE OF 950-1466 FROM N.A.
 RX MEDLINE=88169827; PubMed=3357782;
 RA Mankoo B.S., Daigleish R.;
 RT "Human pro alpha 1(III) collagen: cDNA sequence for the 3' end.";
 RL Nucleic Acids Res. 16:2337-2337(1988).
 RN [8].
 RP REVISION TO 1184.
 RX MEDLINE=89058346; PubMed=3211760;
 RA Molineux K., Daigleish R.;
 RT "Human type III collagen 'variant' is a cDNA cloning artefact.";
 RL Nucleic Acids Res. 16:11833-11833(1988).
 RN [9].
 RP SIMQUENCE OF 1065-1466 FROM N.A.
 RX MEDLINE=85087944; PubMed=6096827;
 RA Leidal H.R., Brinker J.M., May M., Pihlajaniemi T., Morrow S.,
 RA Rosenblom J., Myers J.C.;
 RT "Molecular cloning and carboxyl-propeptide analysis of human type III procollagen.";
 RL Nucleic Acids Res. 12:9383-9394(1984).
 RN [10].
 RP SIMQUENCE OF 965-1200.
 RX MEDLINE=81208139; PubMed=7016180;

RA Seyer J.M., Kang A.H.;
 RT "Covalent structure of collagen: amino acid sequence of alpha
 RT I (III)-CB9 from type III collagen of human liver.";
 RL Biochemistry 20:2621-2627(1981).
 RN [11]

RP SEQUENCE OF 1176-1466 FROM N.A.
 RX MEDLINE=8515760; PubMed=2579949;
 RA Chu M.-L., Weil D., de Wet W.J., Bernard M.P., Sippola M., Ramirez F.;
 RT "Isolation of cDNA and genomic clones encoding human pro-alpha¹₁ (III) collagen. Partial characterization of the 3' end region of the gene.";
 RT J. Biol. Chem. 260:4357-4363(1985).
 RL [12]

RN [12]

RP SEQUENCE OF 1161-1200 FROM N.A.
 RX MEDLINE=8618704; PubMed=3754462;

RA Miskulin M., Dalgleish R., Kluwe-Beckerman B., Rennard S.I.,
 RA Tolstoshev P., Brantly M., Crystal R.G.;
 RT "Human type III collagen gene expression is coordinately modulated
 with the type I collagen genes during fibroblast growth.";
 RL Biochemistry 25:1408-1413(1986).
 RN [13]

RC SEQUENCE OF 1-170 FROM N.A.
 RX TISSUE=Placenta;
 MEDLINE=8303360; PubMed=3405773;
 RA Toman D., Ricca G., de Crombrugge B.;
 RT "Nucleotide sequence of a cDNA coding for the amino-terminal region
 of human prepro alpha 1(III) collagen.";
 RL Nucleic acids Res. 16:7201-7201(1988).
 RN [14]

RP SEQUENCE OF 1-176 FROM N.A.
 RX MEDLINE=89378752; PubMed=2777083;
 RA Benson-Chanda V., Su M.W., Weil D., Chu M.-L., Ramirez F.;
 RT "Cloning and analysis of the 5' portion of the human type-III
 procollagen gene (COL3A1).";
 RL Gene 78:255-265(1999).
 RN [15]

RP REVIEW ON VARIANTs.
 RX MEDLINE=9725559; PubMed=9101290;
 RA Kuivaniemi H., Tromp G., Prockop D.J.;
 RT "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-
 associated collagen (type IX), and network-forming collagen (type X),
 cause a spectrum of diseases of bone, cartilage, and blood vessels.";
 RL Hum. Mutat. 9:300-315(1997).
 RN [16]

RP VARIANT AORTIC ANEURYSM ARG-303, AND VARIANT THR-668.
 RX MEDLINE=9329388; PubMed=8514866;

RA Tromp G., Prockop D.J., Madhatheri S.L., Kleinert C., Darling R.C., Abbott W.M.,
 RA Barley J.J., Zhuang J., Noergaard O., Darling R.C., Abbott W.M.,
 RA Cole C.W., Jaakkola P., Ryynenen M., Pearce W.H., Yao J.S.T.,
 RA Majamaa K., Smullen L., Galatica Z., Ferrell R.E., Jimenez S.A.,
 RA Jackson C.E., Michaels V.V., Kaye M., Kuivaniemi H.;
 RT "Sequencing of cDNA from 50 unrelated patients reveals that mutations
 in the triple-helical domain of type III procollagen are an
 infrequent cause of aortic aneurysms.";
 RL J. Clin. Invest. 91:2539-2545(1993).
 RN [17]

RP VARIANT THR-698.
 RX MEDLINE=91045136; PubMed=2235526;

RA Zafarullah K., Kleinhert C., Tromp G., Kuivaniemi H., Kontusaari S.,
 RA Wu Y., Ganguly A., Prockop D.J.;
 RT "G to A polymorphism in exon 31 of the COL3A1 gene.";
 RL Nucleic Acids Res. 18:6180-6180(1990).
 RN [18]

RP VARIANT AORTIC ANEURYSM ARG-786.
 RX MEDLINE=91056145; PubMed=224125;

RA Kontusaari S., Tromp G., Kuivaniemi H., Romanic A.M., Prockop D.J.;
 RT "A mutation in the gene for type III procollagen (COL3A1) in a family
 with aortic aneurysm.";
 RL J. Clin. Invest. 86:1465-1473(1990).
 RN [19]

RP VARIANT EDS-IV ARG-828.
 RX MEDLINE=94016385; PubMed=8411057;
 Richards A.J., Narcisi P., Lloyd J.C., Ferguson C., Pope F.M.;

RT "The substitution of glycine 661 by arginine in type III collagen
 produces mutant molecules with different thermal stabilities and
 causes Ehlers-Danlos syndrome type IV.";
 RL J. Med. Genet. 30:690-693(1993).
 RN [20]

RP VARIANT EDS-IV SER-957.
 RX MEDLINE=89109135; PubMed=2492273;
 RA Chu M.-L., Kluwe-Beckerman B., Shikata H., Prockop D.J.;
 RT "A single base mutation that substitutes serine for glycine 790 of
 the alpha 1 (III) chain of type III procollagen exposes an arginine
 and causes Ehlers-Danlos syndrome type IV.";
 RL J. Biol. Chem. 264:1349-1352(1989).
 RN [21]

RP VARIANT EDS-IV VAL-960.
 RX MEDLINE=95268429; PubMed=779417;

RA Tromp G., de Paep A., Nyuytinck L., Madhatheri S.L., Kuivaniemi H.;
 RT "Substitution of valine for glycine 793 in type III procollagen in
 Ehlers-Danlos syndrome type IV.";
 RL Hum. Mutat. 5:179-181(1995).
 RN [22]

RP VARIANT EDS-IV GLU-1014.
 RX MEDLINE=92316511; PubMed=1352273;
 RA Richards A.J., Ward P.N., Narcisi P., Nicholls A.C., Lloyd J.C.,
 RA Pope F.M.;
 RT "A single base mutation in the gene for type III collagen (COL3A1)
 converts glycine 847 to glutamic acid in a family with Ehlers-Danlos
 syndrome type IV. An unaffected family member is mosaic for the
 mutation.";
 RL Hum. Genet. 89:414-418(1992).
 RN [23]

RP VARIANT EDS-IV ASP-1050.
 RX MEDLINE=90037070; PubMed=2808425;

RA Tromp G., Kuivaniemi H., Stollie C.A., Pope F.M., Prockop D.J.,
 RA Richards A.J., Lloyd J.C., Ward P.N., de Paep A., Narcisi P.,
 RA Pope F.M.;
 RT "Single base mutation in the type III procollagen gene that converts
 the codon for glycine 883 to aspartate in a mild variant of
 Ehlers-Danlos syndrome type IV.";
 RL J. Biol. Chem. 264:19313-19317(1989).
 RN [24]

RP VARIANT EDS-IV VAL-1077.
 RX MEDLINE=91374480; PubMed=1895316;

RA Richards A.J., Lloyd J.C., Ward P.N., de Paep A., Narcisi P.,
 RA Pope F.M.;
 RT "Characterisation of a glycine to valine substitution at amino acid
 position 101 of the triple helical region of type III collagen in a
 patient with Ehlers-Danlos syndrome type IV.";
 RL J. Med. Genet. 28:458-463(1991).
 RN [25]

RP VARIANT EDS-IV GLU-1173.
 RX MEDLINE=93022543; PubMed=1357232;
 RA Johnson P.H., Richards A.J., Pope F.M., Hopkinson D.A.;
 RA Johnson P.H., Richards A.J., Pope F.M.,
 RT "Characterisation of a glycine to valine substitution at amino acid
 position 101 of the triple helical region of type III collagen in a
 patient with Ehlers-Danlos syndrome type IV.";
 RL J. Med. Genet. 28:458-463(1991).

Query Match 65.2%; Score 60; DB 1; Length 1466;
 Best Local Similarity 78.6%; Pred. No. 0.28;
 Matches 11; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GVKGDKGNPGWPGA 14
 ||||| |||||
 Db 738 GPDKDGEPGGPGA 751

RESULT 19
 CA34_BOVIN
 ID CA34_BOVIN
 AC 02B084;
 DT 01-Nov-1997 (Rel. 35, created)
 DT 01-Nov-1997 (Rel. 35, last sequence update)
 DT 01-Nov-1997 (Rel. 35, last annotation update)
 DE Collagen alpha 3(IV) chain (Fragment).
 GN COL43.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoideae;
 Bovidae; Bovinae; Bos.

OX NCBL_TAXID=9913;
 RN [1]
 SEQUENCE FROM N.A.
 RP
 RC TISSUE="tissue";
 RX MEDLINE=91093146; PubMed=1985905;
 RA MORRISON K.E.; Gemino G.G.; Reeder S.T.;
 RT "Use of the polymerase chain reaction to clone and sequence a cDNA
 encoding the bovine alpha 3 chain of type IV collagen.";
 RL J. Biol. Chem. 266:34-39 (1991).
 J. Biol. Chem. 265:5466-5469(1990).
 [2]
 SEQUENCE OF 227-258.
 RC TISSUE=kidney;
 RX MEDLINE=90202779; PubMed=2318822;
 RA Gunwar S.; Saus J.; Noeiken M.E.; Hudson B.G.;
 RT "Glomerular basement membrane. Identification of a fourth chain,
 alpha 4, of type IV collagen.";
 RL J. Biol. Chem. 265:5466-5469(1990).
 RN [3]
 SEQUENCE OF 227-254.
 RX MEDLINE=83330844; PubMed=3417661;
 RA Saus J.; Wieslander J.; Langeveld J.P.M.; Quinones S.; Hudson B.G.;
 RT "Identification of the Goodpasture antigen as the alpha 3(IV) chain
 of collagen IV.";
 RL J. Biol. Chem. 263:13374-13380(1988).
 RN [4]
 SEQUENCE OF 227-244.
 RX MEDLINE=8722419; PubMed=2438283;
 RA Butkowski R.J.; Langeveld J.P.M.; Wieslander J.; Hamilton J.;
 RA Hudson B.G.;
 RT Localization of the goodpasture epitope to a novel chain of basement
 membrane collagen.";
 J. Biol. Chem. 262:7874-7877(1987).
 RL CC
 CC -I- FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF
 GLOMERULAR BASEMENT MEMBRANES (GBM), FORMING A 'CHICKEN-WIRE'
 MESHWORK TOGETHER WITH LAMININS, PROTEOGLYCANS AND ENACTIN/
 CC NIDOCIN.
 CC -I- SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOPORMS, ALPHA 1(IV)-
 ALPHA 6(IV), EACH OF WHICH CAN FORM A TRIPLE HELIX STRUCTURE
 WITH 2 OTHER CHAINS TO GENERATE TYPE IV COLLAGEN NETWORK.
 CC -I- SUBCELLULAR LOCATION: CELL SURFACE (POSSIBLY).
 CC -I- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS
 DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE
 G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY
 CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL
 TRIPLE-HELICAL 7S DOMAIN.
 CC -I- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPETIDE REPEATING
 UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 CC -I- PTM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH
 ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF
 THESE, LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE
 IV COLLAGENS.
 CC -I- SIMILARITY: TO OTHER TYPE IV COLLAGENS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license/agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 DR EMBL; M63139; AAAK62708; 1; -.
 DR InterPro; IPR0142; C4.
 DR InterPro; IPR000087; Collagen.
 DR Pfam; PF01413; C4; 2.
 DR Pfam; PF01391; Collagen; 4.
 DR Pfam; PF00592; Collagen.
 DR Pfam; PF00593; Collagen.
 DR Pfam; PF00594; Collagen.
 DR Pfam; PF00595; Collagen.
 DR Pfam; PF00596; Collagen.
 DR Pfam; PF00597; Collagen.
 DR Pfam; PF00598; Collagen.
 DR Pfam; PF00599; Collagen.
 DR SMART; SM00111; C4; 2.
 DR Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 KW Glycoprotein; Basement membrane; Collagen; Cell adhesion.
 FT NON-TER 1
 FT DOMAIN <1 238 TRIPLE-HELICAL REGION.
 FT DOMAIN 239 471 NONHELICAL REGION (NC1).

RESULT 20
 CA14_CABEL STANDARD: PRP; 1758 AA.
 ID CA14_CABEL
 AC PA7139;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-MAR-1992 (Rel. 21, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Collagen alpha 1(IV) chain precursor.
 RN EMB-9 OR Clia-2 OR KO414_1.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;
 OC Rhabditidae; Plectoderaiae; Caenorhabditidae;
 OC NCBI_TAXID6239;
 RN [1]
 SEQUENCE FROM N.A.
 RP STRAIN=BRISTOL N2;
 RX MEDLINE=91141582; PubMed=1996137;
 RA Guo X.; Johnson J.J.; Kramer J.M.;
 RT "Embryonic lethality caused by mutations in basement membrane
 collagen of *C. elegans*";
 RL Nature 349:707-709(1991).
 RN [2]
 SEQUENCE FROM N.A.
 RP STRAIN=BRISTOL N2;
 RX MEDLINE=9150718; PubMed=7906398;
 RA Wilson R.; Alnscough R.; Anderson K.; Baynes C.; Berks M.; Coulson A.;
 RA Bonfield J.; Burton J.; Connell M.; Copsey T.; Cooper J.; Dear S.; Du Z.; Durbin R.; Favell A.; Fraser A.;
 RA Craxton M.; Dear S.; Du Z.; Durbin R.; Favell A.; Fraser A.;
 RA Fulton L.; Gardner A.; Green P.; Hawkins T.; Hillier L.; Jier M.;
 RA Johnston L.; Jones M.; Kershaw J.; Kirsten J.; Laister N.;
 RA Latreille P.; Lightning J.; Lloyd C.; Mortimore B.; O'Callaghan M.;
 RA Parsons J.; Percy C.; Rifkin L.; Roopra A.; Saunders D.; Showken R.;
 RA Sims M.; Smalldon N.; Smith A.; Smith M.; Sonnhammer E.; Staden R.;
 RA Sulston J.; Thierry-Mieg J.; Thomas K.; Vaudin M.; Vaughan K.;
 RA Waterston R.; Watson A.; Weinstock L.; Wilkinson-Sproat J.;
 RA Willbman P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of *C.*
 elegans";
 RL Nature 368:332-338(1994).
 RN [3]
 SEQUENCE OF 1446-1758 FROM N.A.
 RP STRAIN=BRISTOL N2;
 RX MEDLINE=9008929; PubMed=2793871;
 RA Guo X.; Kramer J.M.;
 RT "The two *Caenorhabditis elegans* basement membrane (type IV) collagen
 genes are located on separate chromosomes";
 J. Biol. Chem. 261:17574-17582(1986).
 RL CC
 CC -I- FUNCTION: COLLAGEN TYPE IV IS SPECIFIC FOR BASEMENT MEMBRANES.
 CC -I- SUBUNIT: TRIMERS OF TWO ALPHA 1(IV) AND ONE ALPHA 2(IV) CHAIN.
 CC TYPE IV COLLAGEN FORMS A MESH-LIKE NETWORK LINKED THROUGH
 CC INTERNAL MOLECULAR INTERACTIONS BETWEEN 7S DOMAINS AND BETWEEN NC1
 CC DOMAINS.

FT	SITE	106	108	CELL ATTACHMENT SITE (POTENTIAL).
FT	MOD_RES	232	232	HYDROXYLATION.
FT	MOD_RES	238	238	OR 349 (BY SIMILARITY).
FT	DISULFID	261	352	OR 352 (BY SIMILARITY).
FT	DISULFID	294	349	BY SIMILARITY.
FT	DISULFID	306	312	OR 463 (BY SIMILARITY).
FT	DISULFID	371	466	OR 465 (BY SIMILARITY).
FT	DISULFID	405	463	BY SIMILARITY.
FT	DISULFID	417	423	
FT	CONFLICT	253	253	S -> Y (IN REF. 3).
SQ	SEQUENCE	471 AA:	47585 MW:	C03B66F14E7008DE CRC64;
Db	Query Match	64.1%	Score 59;	DB 1;
Db	Best Local Similarity	66.7%	Pred. No. 0.13;	
Db	Matches	10;	Conservative	
Db	Mismatches	2;	Indels	0;
Db	Gaps	0;		
Oy	1	GVVKDKGNNGWPGAP 15		
116 GMGKKGNGNFGPP 130				

Query Match	64.1%	Score 59;	DB 1;	Length 471;
Best Local Similarity	66.7%	Pred. No. 0.13;		
Matches	10;	Conservative		
Mismatches	2;	Indels	0;	
Gaps	0;			

Query Match 64.1%; Score 59; DB 1; Length 1758;
 Best Local Similarity 73.3%; Pred. No. 0.47;
 Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 NCBI_TaxID:60606; OX

RT	collagen and tissue-specific differences in the expression of the corresponding transcripts."	CC	detachment; macular abnormalities and occipital encephalocele.
RL	corresponding transcripts."	CC	- - SIMILARITY: BELONGS TO THE FIBRIL-ASSOCIATED COLLAGENS WITH INTERRUPTED HELICES (FACIT) FAMILY.
RN	[2]	CC	-----
RP	SEQUENCE FROM N.A.	CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
RX	MEDLINE=20289799; PubMed=10830953;	CC	-----
RA	Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,	DR	the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
RA	Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Soeda E.,	EMBL	-----
RA	Oki M., Takagi T., Sakai T., Taudien S., Blechschmidt K., Polley A.,	EMBL	AF018081; AAC39558.1; -.
RA	Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,	EMBL	AF018082; AAC39559.1; -.
RA	Reichwald K., Rump A., Schilhabal M., Schudy A., Zimmermann W.,	EMBL	AL163302; CAB90882.1; -.
RA	Rosenthal A., Kudo J., Shibuya K., Kawasaki S.,	EMBL	L2258; AA5164.1; -.
RA	Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,	EMBL	AAF18406; AAFO1310.1; ALT-INIT.
RA	Minoshima S., Shimizu N., Nordstek G., Hornischer K., Brandt P.,	DR	HSRP; P39061; I-KOEV.
RA	Scharfe M., Schoen A., Desario C., Reichelt J., Kauer G., Blocker H.,	DR	GlycoSuiteDB; P39060; -.
RA	Ramser J., Beck A., Klages S., Hennig S., Risselmann L., Dagand E.,	DR	MIM: 120328; -.
RA	Wehmeyer S., Borzym K., Gatzdiner K., Nizetic D., Francis F., Lehrach H., Reinhardt R., Yaspo M.-L.;	DR	InterPro; IPR00087; Collagen.
RT	"The DNA sequence of human chromosome 21.";	DR	InterPro; IPR00179; Laminin_G.
RL	Nature 405:311-319(2000).	DR	PFam; PF01391; Collagen_6.
RN	[3]	DR	InterPro; IPR00210; TSPN_1.
RP	SEQUENCE OF 834-1516 FROM N.A.	DR	SMART; SM0210; TSPN_1.
RX	MEDLINE=94245237; PubMed=8108291;	DR	Extracellular matrix; Connective tissue; Repeat; Hydroxylation; Cell adhesion; Collagen; Glycoprotein; Signal; Alternative splicing; KW
RC	TISSUE=Placenta;	FT	KW POLYMORPHISM.
RA	Zhi-Yong H., Biao L., Wei-Jie Z., Xiang-Fu W.;	FT	SIGNAL. 1 23 POTENTIAL.
RT	"Cloning and expression of human endostatin gene in Escherichia coli.;"	FT	CHAIN 24 1516 COLLAGEN ALPHA 1(XVIII) CHAIN.
RT	Submitted (SEP-1999) to the EMBL/Genbank/DDBJ databases.	FT	CHAIN 1334 1516 ENDOSTATIN.
RN	[5]	FT	DOMAIN 24 516 NONHELCICAL REGION 1 (NC1).
RP	INVOLVEMENT IN KNOBLOCH SYNDROME.	FT	DOMAIN 517 550 TRIPLE-HELICAL REGION 1 (COL1).
RX	MEDLINE=20400145; PubMed=10542434;	FT	DOMAIN 551 560 NONHELCICAL REGION 2 (NC2).
RA	Sertie A.L., Sossi V., Camargo A.A., Zatz M., Brahe C.,	FT	DOMAIN 561 640 TRIPLE-HELICAL REGION 2 (COL2).
RA	Passos-Bueno M.R.;	FT	DOMAIN 641 664 NONHELCICAL REGION 3 (NC3).
RT	"Collagen XVIII, containing an endogenous inhibitor of angiogenesis and tumor growth, plays a critical role in the maintenance of retinal structure and in neural tube closure.";	FT	DOMAIN 665 786 TRIPLE HELICAL REGION 3 (COL3).
RL	Hum. Mol. Genet. 9:2051-2058(2000).	FT	DOMAIN 787 809 NONHELCICAL REGION 4 (NC4).
RN	[6]	FT	DOMAIN 810 892 TRIPLE-HELICAL REGION 4 (COL4).
RP	VARIANT ASN-1437.	FT	DOMAIN 893 906 NONHELCICAL REGION 5 (NC5).
RX	PUBMED=11606364;	FT	DOMAIN 907 948 TRIPLE-HELICAL REGION 5 (COL5).
RA	Iughetti P., Suzuki O., Godoi P.H., Alves V.A., Sertie A.L.,	FT	DOMAIN 949 961 NONHELCICAL REGION 6 (NC6).
RA	Zotick T., Soares F., Camardo A., Moreira E.S., di Loreto C.,	FT	DOMAIN 962 1034 TRIPLE-HELICAL REGION 6 (COL6).
RA	Moreira-Filho C.A., Simpson A., Oliva G., Passos-Bueno M.R.;	FT	DOMAIN 1035 1044 NONHELCICAL REGION 7 (NC7).
RT	"A polymorphism in endostatin, an angiogenesis inhibitor, predisposes for the development of prosstatic adenocarcinoma.";	FT	DOMAIN 1045 1077 TRIPLE-HELICAL REGION 7 (COL7).
RL	Cancer Res. 61:7315-7318(2001).	FT	DOMAIN 1078 1089 NONHELCICAL REGION 8 (NC8).
CC	-- FUNCTION: COLA18A PROBABLY PLAYS A MAJOR ROLE IN DETERMINING THE RETINAL STRUCTURE AS WELL AS IN THE CLOSURE OF THE NEURAL TUBE.	FT	DOMAIN 1109 1111 TRIPLE-HELICAL REGION 8 (COL8).
CC	-- FUNCTION: ENDOSTATIN POTENTLY INHIBITS ENDOTHELIAL CELL PROLIFERATION AND ANGIOGENESIS. MAY INHIBIT ANGIOGENESIS BY BINDING TO THE HEPARAN SULPHATE PROTEOGLYCANS INVOLVED IN GROWTH FACTOR SIGNALLING.	FT	DOMAIN 1112 1118 NONHELCICAL REGION 9 (NC9).
CC	-- ALTERNATIVE PRODUCTS: 2 ISOFORMS: A SHORT FORM/NC1-303 AND A LONG FORM/NC493 (SHOWN HERE); ARE PRODUCED BY ALTERNATIVE SPlicing.	FT	DOMAIN 1119 1173 TRIPLE-HELICAL REGION 9 (COL9).
CC	-- TISSUE SPECIFICITY: PRESENT IN MULTIPLE ORGANS WITH HIGHEST LEVELS IN LIVER, LUNG AND KIDNEY.	FT	DOMAIN 1174 1186 NONHELCICAL REGION 10 (NC10).
CC	-- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPETIDE REPEATING UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.	FT	DOMAIN 1187 1204 TRIPLE-HELICAL REGION 10 (COL10).
CC	-- POLYMORPHISM: There is an association between a polymorphism in position 1437 and prostate cancer. Heterozygous Asn-1437 individuals have a 2.5 times increased chance of developing prostate cancer as compared with homozygous Asp-1437 individuals.	FT	DOMAIN 1205 1516 NONHELCICAL REGION 11 (NC11).
CC	-- DISEASE: Defects in COL18A1 are a cause of Knobloch syndrome (KS or KNO); an autosomal recessive disorder defined by the occurrence of high myopia, vitreoretinal degeneration with retinal	FT	DOMAIN 168 68 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC	[7]	FT	DOMAIN 229 129 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC	INDIVIDUALS: Individuals have a 2.5 times increased chance of developing prostate cancer as compared with homozygous Asp-1437 individuals.	FT	DOMAIN 164 164 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC	DEFECTS: Defects in COL18A1 are a cause of Knobloch syndrome (KS or KNO); an autosomal recessive disorder defined by the occurrence	FT	DOMAIN 691 691 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC	of high myopia, vitreoretinal degeneration with retinal	FT	DOMAIN 1329 1329 O-LINKED (GALNAc. . .) (POTENTIAL).
CC		FT	/FRID=CAR_000150.
CC		FT	BY SIMILARITY.
CC		FT	CELL ATTACHMENT SITE (POTENTIAL).
CC		FT	SITE (POTENTIAL).
CC		FT	MISSING (IN SHORT ISOFORM).
CC		FT	PROPPWPWRRRLDVLAPLVLLGVRAASAEP (IN SHORT ISOFORM).
CC		FT	D->N.
CC		FT	/FRID=VAR_012709.
CC		FT	F->S (IN REF. 2).
CC		FT	I->V (IN REF. 2).
CC		FT	V->L (IN REF. 3).

FT CONFLICT 886 886 G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY CAUSE FLEXIBILITY IN THE TRIPLE HELIX) AND A SHORT N-TERMINAL CC FT CONFLICT 912 912 P -> R (IN REF. 3).
 FT CONFLICT 933 933 R -> R (IN REF. 3).
 FT CONFLICT 975 975 P -> L (IN REF. 3).
 FT CONFLICT 1064 1064 A -> A (IN REF. 3).
 FT CONFLICT 1084 1084 L -> K (IN REF. 3).
 FT CONFLICT 1120 1120 P -> A (IN REF. 3).
 FT CONFLICT 1123 1123 P -> A (IN REF. 3).
 FT CONFLICT 1126 1126 P -> PPGP (IN REF. 2).
 FT CONFLICT 1206 1206 G -> GO (IN REF. 3).
 FT CONFLICT 1304 1304 R -> G (IN REF. 3).
 FT CONFLICT 1314 1314 A -> G (IN REF. 3).
 FT CONFLICT 1324 1324 LR -> CG (IN REF. 3).
 FT CONFLICT 1443 1443 R -> T (IN REF. 4).
 FT CONFLICT 1483 1483 S -> Y (IN REF. 4).
 FT CONFLICT 1516 AA; 153840 MW; 3C70F29A44765B876 CRC64;
 SQ SEQUENCE 1516 AA; 153840 MW; 3C70F29A44765B876 CRC64;

Query Match 63.0%; Score 58; DB 1; Length 1516;
 Best Local Similarity 66.7%; Pred. No. 0; 57; Mismatches 1; Indels 4; Gaps 0; DR
 Matches 10; Conservative 1; MisMatches 4; Indels 0; Gaps 0; DR
 Qy 1 GVKGDKGNPGNGAP 15 DR
 Db 517 GLKGQKGEPPGVPGPP 531 DR

RESULT 23
CA24_CAEEL STANDARD; PRT; 1758 AA.
 ID CA24_CAEEL
 AC P17140;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 30-MAR-2000 (Rel. 39, Last annotation update)
 DE Collagen alpha 2(IV) chain precursor.
 GN LET-2 OR CBL-1.
 OS Caenorhabditis elegans.
 OC Rhabditidae; Plecteridae; Caenorhabditis; Rhabditida; Rhabditoidea;
 OC Rhabditida; Rhabditidae; Caenorhabditis; Rhabditida; Rhabditoidea;
 OC NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N A.
 RC STRAIN-BRISTOL_N2;
 RX MEDLINE="94012964"; PubMed="7691828;
 RR Sibley M.H., Johnson J.J., Mello C.C., Kramer J.M.;
 RT "Genetic identification, sequence, and alternative splicing of the
 Caenorhabditis elegans alpha 2(IV) collagen gene.",
 RL J. Cell Biol. 123:255-264(1993).
 RN [2]
 RP PRELIMINARY SEQUENCE OF 1495-1758 FROM N A.
 RC STRAIN-BRISTOL_N2;
 RX MEDLINE="90089129"; PubMed="2793871;
 RA Guo X., Kramer J.M.;
 RT "The two Caenorhabditis elegans basement membrane (type IV) collagen genes are located on separate chromosomes.",
 RL J. Biol. Chem. 264:17574-17582(1989).
 RN [3]
 RP VARIANTS.
 RX MEDLINE="94320591"; PubMed="8045258;
 RA Sibley M.H., Graham P.L., von Mende N., Kramer J.M.;
 RT "Mutations in the alpha 2(IV) basement membrane collagen gene of Caenorhabditis elegans produce phenotypes of differing severities.",
 RL EMBO J. 13:3278-3285(1994).
 CC -I- FUNCTION: COLLAGEN TYPE IV IS SPECIFIC FOR BASEMENT MEMBRANES.
 CC -I- SUBUNIT: TRIMERS OF TWO ALPHA 1(IV) AND ONE ALPHA 2(IV) CHAIN.
 CC TYPE IV COLLAGEN FORMS A MESH-LIKE NETWORK LINKED THROUGH
 CC INTERMOLECULAR INTERACTIONS BETWEEN 7S DOMAINS AND BETWEEN NC1
 CC DOMAINS.
 CC -I- ALTERNATIVE PRODUCTS: 2 ISOFORMS: I (SHOWN HERE) AND II; ARE
 CC PRODUCED BY ALTERNATIVE SPlicing.
 CC -I- TISSUE SPECIFICITY: FORM I IS PREDOMINANT IN EMBRYOS AND FORM
 CC II IS PREDOMINANT IN THE LARVAE AND ADULTS.
 CC -I- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS
 CC DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE
 CC LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE
 CC IV COLLAGENS.
 CC -I- DISEASE: MUTATIONS IN LET-2 ARE GENERALLY EMBRYONIC LETHAL.
 CC -I- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPETIDE REPEATING
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 CC -I- PTM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH
 CC ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF
 CC THESE, LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE
 CC IV COLLAGENS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>)
 CC or send an email to license@isb-sib.ch).

DR EMBL; 222954; CAR80536; 1;
 DR EMBL; 222964; CAR80537; 1;
 DR PR0D0M; PD003923; C4; 2.
 DR SMART; SM0111; C4; 2.
 DR PIR; A34476; A34476.
 DR InterPro; IPR001442; C4.
 DR InterPro; IPR000887; Collagen.
 DR Pfam; PF01413; C4; 2.
 DR Pfam; PF01391; Collagen; 23.
 DR PR0D0M; PD003923; C4; 2.
 DR SMART; SM0111; C4; 2.
 KW Hydroxylation; Connective tissue; Basement membrane; Repeat; Collagen;
 KW Alternative splicing; Glycoprotein; Signal.
 FT SIGNAL 1 26 POTENTIAL.
 FT CHAIN 27 1758 COLLAGEN ALPHABET 2(IV) CHAIN.
 FT DOMAIN 27 1758 7S DOMAIN.
 FT DOMAIN 42 1527 TRIPLE-HELICAL REGION.
 FT DOMAIN 1528 1758 NONHELIICAL REGION (NC1).
 FT DISULFID 1546 1635 DISULFID OR 1632 (BY SIMILARITY).
 FT DISULFID 1579 1635 OR 1635 (BY SIMILARITY).
 FT DISULFID 1591 1597 BY SIMILARITY.
 FT DISULFID 1654 1750 OR 1747 (BY SIMILARITY).
 FT DISULFID 1688 1747 OR 1750 (BY SIMILARITY).
 FT DISULFID 1700 1707 BY SIMILARITY.
 FT CARBOHD 248 248 O-LINKED (GLUCOSAMINOGLYCAN) (POTENTIAL).
 FT VARSPLIC 229 264 DDGAGSPGGVPGPVGPPGPPIASTMSKGTTGPKGDLGEGEK (IN
 FT ISOFORM II).
 FT VARIANT 48 48 G -> E (IN MN14; 73% LETHAL).
 FT VARIANT 366 366 A -> T (IN MN126; 100% LETHAL).
 FT VARIANT 570 570 G -> E (IN MN12; 37% LETHAL).
 FT VARIANT 588 588 G -> R (IN MN103 AND MN151; 96% LETHAL).
 FT VARIANT 597 597 G -> R (IN MN152; 50% LETHAL).
 FT VARIANT 690 690 G -> R (IN MN101; 100% LETHAL).
 FT VARIANT 690 690 G -> E (IN MN129; 100% LETHAL).
 FT VARIANT 737 737 G -> E (IN MN143; 100% LETHAL).
 FT VARIANT 877 877 G -> R (IN G30; 90% LETHAL).
 FT VARIANT 904 904 G -> R (IN EL470; 94% LETHAL).
 FT VARIANT 1003 1003 G -> E (IN MN139; 20% LETHAL).
 FT VARIANT 1125 1125 G -> D (IN G25; 2% LETHAL).
 FT VARIANT 1152 1152 G -> D (IN MN147; 7% LETHAL).
 FT VARIANT 1286 1286 G -> D (IN G37 AND B246; 9% LETHAL).
 FT VARIANT 1682 1682 P -> L (IN REF. 1).
 FT CONFFLICT 1758 AA; 167750 MW; 97EB3F3DBB2D2AC5 CRC64;
 SQ SEQUENCE 1758 AA; 167750 MW; 97EB3F3DBB2D2AC5 CRC64;

Query Match 63.0%; Score 58; DB 1; Length 1758;
 Best Local Similarity 73.3%; Pred. No. 0; 66; Mismatches 1; Indels 3; Gaps 0; DR
 Matches 11; Conservative 1; MisMatches 3; Indels 0; Gaps 0; DR
 Qy 1 GVKGDKGNPGNGAP 15 DR
 Db 416 GBKGDKGDGGIPGP 430 DR

RESULT 24

CC24_CASEL	STANDARD;	PRT;	298 AA.
ID .CC24_CASEL	RN [1]		
AC P3687;	RP		
DT 01-FEB-1994 (Rel. 28, Created)	SEQUENCE FROM N.A.		
DT 01-FEB-1994 (Rel. 28, Last sequence update)	MEDLINE-9242649; PubMed=10224290;		
DT 01-JUN-1994 (Rel. 29, Last annotation update)	RA Suzuki H., Kodama T., Tryggrason K., Roziel H.; Kobzik L.; Palecanda A.; Paulauskis J.; Al-Mutairi E.; Imrich A.; Qin G.,		
DE Cuticle collagen 34.	RT "Role of the scavenger receptor MARCO in alveolar macrophage binding of unpolymerized environmental particles.";		
GN COL-34.	J. Exp. Med. 189:497-1506(1999).		
OS Caenorhabditis elegans.	RL		
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea; Rhabditidae; Peloderrina; Caenorhabditis.	CC -1 - FUNCTION: Bind Gram-positive and Gram-negative bacteria.		
OC NCBI_TaxID=6239;	CC -1 - SUBCELLULAR LOCATION: Type II membrane protein (Probable).		
RN [1]	CC -1 - SIMILARITY: CONTAINS 1 SRCR DOMAIN.		
RP SEQUENCE FROM N.A.	CC		
RC STRAIN=RISTOL N2;	CC		
RX MEDLINE=93013043; PubMed=1398138;	CC		
RA B1d D.M.;	CC		
RT "Sequence comparison of the <i>Caenorhabditis elegans</i> dpy-13 and col-34 genes, and their deduced collagen products";	CC		
RL Gene 120:261-266(1992).	CC		
CC -1 - FUNCTION: NEURODODE CUTICLES ARE COMPOSED largely OF COLLAGEN-LIKE PROTEINS. THE CUTICLE FUNCTIONS BOTH AS AN EXOSKELETON AND AS A BARRIER TO PROTECT THE NERM FROM ITS ENVIRONMENT.	CC		
CC -1 - SUBUNIT: COLLAGEN POLYPEPTIDE CHAINS ARE COMPLEXED WITHIN THE CUTICLE BY DISULFIDE BONDS AND OTHER TYPES OF COVALENT CROSS-LINKS.	CC		
CC -1 - SIMILARITY: TO OTHER COLLAGENS. STRONG, TO OTHER CUTICLE COLLAGENS.	CC		
CC	CC		
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).	CC		
CC	CC		
CC DR EMBL; AF125191; AND20360; 1.	DR		
CC DR InterPro; IPRO0087; Collagen.	DR		
CC DR InterPro; IPRO01190; SRCR.	DR		
CC DR Pfam; PF01391; Collagen; 4.	DR		
CC DR PRINTS; PR00530; SRCR; 1.	DR		
CC DR SMART; SM00202; SPERACTRCPTR.	DR		
CC DR PROSITE; PS00420; SRCR; 1; FALSE_NEG.	DR		
CC DR PROSITE; PS50387; SRCR; 2; 1.	DR		
CC KW Collagen; Transmembrane; Receptor; Glycoprotein; Signal-anchor.	KW		
CC FT DOMAIN 1 48	FT		
CC FT TRANSMEM 49 69	FT		
CC FT DOMAIN 70 483	FT		
CC FT DOMAIN 148 383	FT		
CC FT DOMAIN 389 483	FT		
CC FT DISULFID 412 472	FT		
CC FT DISULFID 425 482	FT		
CC FT DISULFID 452 462	FT		
CC FT CARBOHYD 85 85	FT		
CC FT CARBOHYD 137 137	FT		
CC SQ SEQUENCE 483 AA; C38F18C46505FB6 CRC64;	SQ		
Query Match 62.0%; Score 57; DB 1; Length 483;	QY 1 GVKGDKGNPGWPGAP 15		
Best Local Similarity 73.3%; Pred. No. 0.16; Matches 11; Conservative 0; Mismatches 4; Indels 0; Gaps 0;	Db 260 GVKGDGQGQGLQGV 274		
RESULT 25	RESULT 26		
MRCO_MESAU STANDARD; PRT; 483 AA.	CA21_ONONY		
ID MRCO_MESAU STANDARD; PRT; 483 AA.	CA21_ONCMY		
AC 09WUB9;	STANDARD;	PRT;	1356 AA.
DT 16-OCT-2001 (Rel. 40, Created)	DT 093384;		
DT 16-OCT-2001 (Rel. 40, Last sequence update)	DT 16-OCT-2001 (Rel. 40, Last sequence update)		
DE Macrophage receptor MARCO (Macrophage receptor with collagenous structure).	DE ColIgA2.		
GN MARCO.	GN Oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri).		
OS Mesocricetus auratus (Golden hamster).	OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.		
OC Eukaryota; Mammalia; Eutheria; Rodentia; Sciurognathii; Muridae; Cricetinae;	OC NCBI_TaxID=8022;		
OC	RN [1]		
RP SEQUENCE FROM N.A.	RP SEQUENCE FROM N.A.		
RX MEDLINE=21257802; PubMed=11358497;	RX Saito M., Takenouchi Y., Kunisaki N., Kimura S.;		
RA	RA		

*Complete primary structure of rainbow trout type I collagen consisting of alpha1(I)-alpha2(I)-alpha3(I) heterotrimeres.;
 RT Kimura S., Kunisaki N., Hiroto I., Aoki T., Ishida M., Urano N., RA
 RT Saito M., Saito M., Kunisaki N., Hiroto I., Aoki T., Ishida M., Urano N., RA
 RT Partial characterization of cDNA clones encoding the three distinct RT
 RT protein chains of type I collagen from rainbow trout.;
 RL Fisheries Sci. 64:780-786(1998).
 CC -!- FUNCTION: TYPE I COLLAGEN IS A MEMBER OF GROUP I COLLAGEN
 CC -!- (FIBRILLAR FORMING COLLAGEN).
 CC -!- SUBUNIT: TRIMERS OF ONE ALPHA 2(I) AND TWO ALPHA 1(I) CHAINS.
 CC -!- TISSUE SPECIFICITY: FORMS THE FIBRILLS OF TENDON, LIGAMENTS AND
 CC BONES. IN BONES THE FIBRILLS ARE MINERALIZED WITH CALCIUM
 CC HYDROXYAPATITE.
 CC -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPETIDE REPEATING
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; AB052637; BAB55663.1; .
 DR EMBL; AB008372; BAA33379.1; .
 DR InterPro; IPR000887; Collagen.
 DR InterPro; IPR00885; Fib_Collagen_C.
 DR InterPro; IPR01410; Colf1; 1.
 DR InterPro; IPR02078; Fib_Collagen_C; 1.
 DR InterPro; SM00388; Colf1; 1.
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 KW Glycoprotein; Collagen; Signal.
 FT SIGNAL 1 24 POTENTIAL.
 FT PROPEP 25 ? AMINO-TERMINAL PROPEPTIDE (POTENTIAL).
 FT CHAIN ? 1096 COLLAGEN ALPHA 2(I) CHAIN.
 FT PROPEP 1097 1356 CARBOXY-TERMINAL PROPEPTIDE
 FT (BY SIMILARITY).
 FT CARBOHYD 1257 1257 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 1356 AA; 126985 MW; 7BB2P1F80DBB10C93 CRC64;
 Query Match 62.0%; Score 57; DB 1; Length 1356;
 Best Local Similarity 71.4%; Pred. No. 0.71; SR. 1;
 Matches 10; Conservative 1; Mismatches 3; Indels 0; Gaps 0; DR
 QY 1 GVKGDKGNPGWGA 14 PROSITE; PS00420; SRCR 1; FALSE_NEG.
 Db 937 GFKGDRGPSPGA 950 PROSITE; PS00287; SRCR 2; 1.
 DR
 DR Pftam; PR00530; SRCR; 1.
 DR PRINS; PR00258; SFERACTRCPR.
 DR SMART; SM00202; SR. 1.
 DR InterPro; IPR00087; Marco.
 DR InterPro; IPR01190; SRCR.
 DR Pftam; PR01391; Collagen; 4.
 DR Pftam; PR00530; SRCR; 1.
 DR PRINS; PR00258; SFERACTRCPR.
 DR SMART; SM00202; SR. 1.
 DR PROSITE; PS00420; SRCR 1; FALSE_NEG.
 DR PROSITE; PS00287; SRCR 2; 1.
 DR
 DR Collagen; Transmembrane; Receptor; Glycoprotein; Signal-anchor.
 DR DOMAIN 1 48 CYTOSLAMIC (POTENTIAL).
 DR TRANSMEM 49 69 SIGNAL ANCHOR (TYPE-II MEMBRANE PROTEIN)
 DR DOMAIN 70 518 (POTENTIAL).
 DR DOMAIN 149 418 EXTRACELLULAR (POTENTIAL).
 DR DOMAIN 423 518 COLLAGEN-LIKE.
 DR DISULFID 446 507 SRCR.
 DR DISULFID 459 517 BY SIMILARITY.
 DR DISULFID 487 497 BY SIMILARITY.
 DR CARBOHYD 87 138 N-LINKED (GLCNAC. . .) (POTENTIAL).
 DR CARBOHYD 138 138 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 518 AA; 52730 MW; B09E7601ECA23637 CRC64;
 Query Match 60.9%; Score 56; DB 1; Length 518;
 Best Local Similarity 66.7%; Pred. No. 0.39; SR. 1;
 Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0; DR
 QY 1 GVKGDKGNPGWGA 15 PROSITE; PS00420; SRCR 1; FALSE_NEG.
 Db 294 GVKGDRGPSPGA 308 PROSITE; PS00287; SRCR 2; 1.
 DR
 DR RESULT 28

CALI_HUMAN
 CALI_HUMAN STANDARD; PRM; 1143 AA.
 ID 01493; 013676; 012885; Q9H572; Q9NQ22; Q9NP22;
 AC DT 01-MAR-2002 (Rel. 41, Created)
 DT 01-MAR-2002 (Rel. 41, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Collagen alpha 1(XIX) chain precursor (Collagen alpha 1(Y) chain).
 GN COL19A1.
 OC Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OC NCBI_TaxID:9606;
 RN [1] SEQUENCE FROM N.A.
 RP MEDLINE:9523914; Pubmed=7775380;
 RA Inoguchi K., Yoshioka H., Khaleduzzaman M., Ninomiya Y.;
 RT "The mRNA for alpha 1(XIX) collagen chain, a new member of FACITs,
 RT contains a long unusual 3' untranslated region and displays many
 RT unique splicing variants.";
 J. Biochem. 117:137-146(1995).
 RN [2] SEQUENCE FROM N.A.
 RX MEDLINE:98008918; Pubmed=9344653;
 RA Khaleduzzaman M., Sumiyoshi H., Ueki Y., Inoguchi K., Ninomiya Y.,
 RA Yoshioka H.; Structure of the human type XIX collagen (COL19A1) gene, which
 RT suggests it has arisen from an ancestor gene of the FACIT family.;
 RL Genomics 45:304-312(1997).
 RN [3] SEQUENCE OF 132-952 FROM N.A.
 RP BLAKEY S., PARKER A., SKUCE C.;
 RA DION A.S.;
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
 RN [4] SEQUENCE OF 739-1143 FROM N.A.
 RC TISSUE-SKIN;
 RX MEDLINE:94308092; Pubmed=8034603;
 RA Myers J.C., Yang H., D'Ippolito J.A., Presente A., Miller M.K.,
 RA Dion A.S.;
 RT "The triple-helical region of human type XIX collagen consists of
 multiple collagenous subdomains and exhibits limited sequence homology
 to alpha 1(XVI).";
 RT J. Biol. Chem. 269:18549-18557(1994).
 CC -1- FUNCTION: May act as a cross-bridge between fibrils and other
 CC extracellular matrix molecules.
 CC -1- PTM: PROLINE AT THE THIRD POSITION OF THE TRIPETIDE REPEATING
 UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS (BY
 CC SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE FIBRIL-ASSOCIATED COLLAGENS WITH
 CC INTERRUPTED HELICES (FACIT) FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement. (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL; D38163; BAA07368; 1; -.
 DR EMBL; AB004629; BAA23309; 1; -.
 DR EMBL; AB004634; BAA23309; 1; JOINED.
 DR EMBL; AB004630; BAA23309; 1; JOINED.
 DR EMBL; AB004615; BAA23309; 1; JOINED.
 DR EMBL; AB004626; BAA23309; 1; JOINED.
 DR EMBL; AB004630; BAA23309; 1; JOINED.
 DR EMBL; AB004631; BAA23309; 1; JOINED.
 DR EMBL; AB004632; BAA23309; 1; JOINED.
 DR EMBL; AB004633; BAA23309; 1; JOINED.
 DR EMBL; AB004584; BAA23309; 1; JOINED.
 DR EMBL; AB004585; BAA23309; 1; JOINED.
 DR EMBL; AB004586; BAA23309; 1; JOINED.
 DR EMBL; AB004587; BAA23309; 1; JOINED.
 DR EMBL; AB004588; BAA23309; 1; JOINED.

DR EMBL; AB004589; BAA23309; 1; JOINED.
 DR EMBL; AB004590; BAA23309; 1; JOINED.
 DR EMBL; AB004591; BAA23309; 1; JOINED.
 DR EMBL; AB004592; BAA23309; 1; JOINED.
 DR EMBL; AB004593; BAA23309; 1; JOINED.
 DR EMBL; AB004594; BAA23309; 1; JOINED.
 DR EMBL; AB004595; BAA23309; 1; JOINED.
 DR EMBL; AB004596; BAA23309; 1; JOINED.
 DR EMBL; AB004597; BAA23309; 1; JOINED.
 DR EMBL; AB004598; BAA23309; 1; JOINED.
 DR EMBL; AB004599; BAA23309; 1; JOINED.
 DR EMBL; AB004600; BAA23309; 1; JOINED.
 DR EMBL; AB004601; BAA23309; 1; JOINED.
 DR EMBL; AB004602; BAA23309; 1; JOINED.
 DR EMBL; AB004603; BAA23309; 1; JOINED.
 DR EMBL; AB004605; BAA23309; 1; JOINED.
 DR EMBL; AB004606; BAA23309; 1; JOINED.
 DR EMBL; AB004612; BAA23309; 1; JOINED.
 DR EMBL; AB004613; BAA23309; 1; JOINED.
 DR EMBL; AB004614; BAA23309; 1; JOINED.
 DR EMBL; AB004615; BAA23309; 1; JOINED.
 DR EMBL; AB004616; BAA23309; 1; JOINED.
 DR EMBL; AB004617; BAA23309; 1; JOINED.
 DR EMBL; AB004618; BAA23309; 1; JOINED.
 DR EMBL; AB004619; BAA23309; 1; JOINED.
 DR EMBL; AB004620; BAA23309; 1; JOINED.
 DR EMBL; AB004621; BAA23309; 1; JOINED.
 DR EMBL; AB004622; BAA23309; 1; JOINED.
 DR EMBL; AB004623; BAA23309; 1; JOINED.
 DR EMBL; AB004624; BAA23309; 1; JOINED.
 DR EMBL; AB004625; BAA23309; 1; JOINED.
 DR EMBL; AB004627; BAA23309; 1; JOINED.
 DR EMBL; AB00463; CAB9931; 1; -.
 DR EMBL; AL133388; CAC0139; 1; -.
 DR EMBL; AL136445; CAC12699; 1; -.
 DR EMBL; U0929; AM21146; 1; -.
 DR EMBL; U09280; AAA21147; 1; -.
 DR EMBL; U09281; AAA21147; 1; JOINED.
 DR MIM; 120165; -.
 DR Interpro; IPR000087; Collagen.
 DR Interpro; IPR003129; TSPN; 12.
 DR Pfam; PF02210; Collagen; 12.
 DR Pfam; PF02210; TSPN; 1.
 DR SMART; SM00210; TSPN; 1.
 DR KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation; Cell adhesion; Collagen; Signal; Potential.
 FT SIGNAL 1 23 POTENTIAL.
 FT CHAIN 24 1143 COLLAGEN ALPHA 1(XIX) CHAIN.
 FT DOMAIN 50 235 TSPN.
 FT DOMAIN 293 352 TRIPLE-HELICAL REGION 1 (COL1).
 FT DOMAIN 371 430 TRIPLE-HELICAL REGION 2 (COL2).
 FT DOMAIN 469 689 TRIPLE-HELICAL REGION 3 (COL3).
 FT DOMAIN 701 819 TRIPLE-HELICAL REGION 4 (COL4).
 FT DOMAIN 834 1013 TRIPLE-HELICAL REGION 5 (COL5).
 FT DOMAIN 1055 1112 TRIPLE-HELICAL REGION 6 (COL6).
 FT SITE 953 955 CELL ATTACHMENT SITE (POTENTIAL).
 FT CONFLICT 89 90 MY -> I (IN REF. 1).
 FT CONFLICT 280 90 Q -> L (IN REF. 1).
 FT CONFLICT 366 366 D -> V (IN REF. 1).
 FT CONFLICT 442 443 YY -> DD (IN REF. 1).
 FT CONFLICT 938 938 Q -> E (IN REF. 1).
 FT CONFLICT 1141 1141 C -> G (IN REF. 1 AND 4).
 SQ SEQUENCE 1143 AA; 115448 MW; 0F1AO0D825589B378 CRC64;

Query Match 60.9%; Score 56; DB 1; Length 1143;
 Best Local Similarity 66.7%; Pred. No. 0.85; Matches 10; Conservative 1; Mismatches 4; Index 0; Gaps 0;

Db 457 |:||||| | ||| |
 ID CA13_MOUSE STANDARD; PRT; 1464 AA.
 AC P08121; 061429; Q9GRN7;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Collagen alpha 1(III) chain precursor.
 GN COL1A1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C7BL X DBA; TISSUE=Embryo;
 RX MEDLINE=55011609; PubMed=7920795;
 RA Toman D., de Crombrugge B.;
 RT "The mouse type-III procollagen-encoding gene: genomic cloning and
 complete DNA sequence.";
 RT Gene 147:161-168(1994).
 RN [2]
 RP SEQUENCE OF 1-488 FROM N.A.
 RC MEDLINE=81167858; PubMed=3443309;
 RA Wood L., Theriault N., Vogeli G.;
 RT "Complete nucleotide sequence of the N-terminal domains of the murine
 alpha-1 type-III collagen chain.";
 RT Gene 61:225-230(1987).
 RN [3]
 RP SEQUENCE OF 1-28 FROM N.A.
 RX MEDLINE=5131189; PubMed=3972847;
 RA Liu G., Mudryj M., de Crombrugge B.;
 RT "Identification of the promoter and first exon of the mouse alpha 1
 (III) collagen gene.;"
 RL J. Biol. Chem. 260:3773-3777(1985).
 RN [4]
 RP SEQUENCE OF 810-1464 FROM N.A.
 RC STRAIN=C7BL/6J; TISSUE=Embryonic head;
 RX MEDLINE=3108560; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shihata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamamoto I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiba H.,
 RA Kuehl P., Lewis S., Matsushige Y., Nikaido T., Pesole G., Quackenbush J.,
 RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Rono H., Baldarelli R., Baish G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Garibaldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato Y., Schoenbach C., Sepe T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsuki S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RN [5]
 RP SEQUENCE OF 1442-1464 FROM N.A.
 RC STRAIN=C7BL;
 RX MEDLINE=61274355; PubMed=2054384;
 RA Mettaerranta M., Toman D., de Crombrugge B., Vuorio E.;
 RT "Specific hybridization probes for mouse type I, II, III and IX
 collagen mRNAs.";
 RL Biochim. Biophys. Acta 1089:241-243(1991).
 CC -!- FUNCTION: COLLAGEN TYPE III OCCURS IN MOST SOFT CONNECTIVE TISSUES
 ALONG WITH TYPE I COLLAGEN.

RESULT 29

CC -!- SUBUNIT: TRIMERS OF IDENTICAL ALPHA 1(III) CHAINS. THE CHAINS ARE
 LINKED TO EACH OTHER BY INTERCHAIN DISULFIDE BONDS. TRIMERS ARE
 ALSO CROSS-LINKED VIA HYDROXYLINES.

CC -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPETIDE REPEATING
 UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.

CC O-LINKED GLYCAN CONSISTS OF GLC-GAL DISACCHARIDE (BY SIMILARITY).

CC -!- SIMILARITY: CONTAINS 1 WFCC DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

CC -----

DR EMBL; X52046; CAA3279_1; -;
 DR EMBL; M18933; AAA31338_1; -;
 DR EMBL; K03037; -; NOT_ANNOTATED_CDS.
 DR EMBL; AK01948; BAB31724_1; -;
 DR EMBL; X51983; CAA41048_1; -;
 DR PIR; A22287; A22287;
 DR PIR; A22353; A22353;
 DR PIR; S16373; S16373;
 DR MGDB; MGI:88453; Col3a1.
 DR InterPro; IPR00087; Collagen.
 DR InterPro; IPR00088; Fib.collagen_C.
 DR Pfam; PF01410; COLFI; 1.
 DR Pfam; PF01391; Collagen; 17.
 DR Prodom; PD002078; Fib.collagen_C; 1.
 DR SMART; SM00214; COLFI; 1.
 DR PROSITE; PS01208; WFCC; 1.
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 KW Glycoprotein; Collagen; Signal.
 FT SIGNAL 1 23 BY SIMILARITY.
 FT PROPER 24 154 AMINO-TERMINAL PROPEPTIDE.
 FT CHAIN 155 1203 COLLAGEN ALPHA 1(III) CHAIN.
 FT PROPER 1204 1464 CARBOXYL-TERMINAL PROPEPTIDE.
 FT DOMAIN 31 90 WFCC.
 FT DOMAIN 155 159 NONHELCICAL REGION (N-TERMINAL).
 FT DOMAIN 170 1195 TRIPLE-HELICAL REGION.
 FT DOMAIN 196 1464 NONHELCICAL REGION (C-TERMINAL).
 FT CARBOHYD 262 262 O-LINKED GAL. . (BY SIMILARITY).
 FT MOD_RES 262 262 HYDROXYLATION (BY SIMILARITY).
 FT DOMAIN 283 283 HYDROXYLATION (BY SIMILARITY).
 FT MOD_RES 859 859 HYDROXYLATION (BY SIMILARITY).
 FT MOD_RES 976 976 HYDROXYLATION (BY SIMILARITY).
 FT MOD_RES 1093 1093 HYDROXYLATION (BY SIMILARITY).
 FT MOD_RES 1105 1105 HYDROXYLATION (BY SIMILARITY).
 FT DISULFID 1195 1195 INTERCHAIN (BY SIMILARITY).
 FT DISULFID 1196 1196 INTERCHAIN (BY SIMILARITY).
 SQ SEQUENCE 1464 AA; 138944 MW; 2104EC27A886090B CRC64;
 Query Match 60.9%; Score 56; DB 1; Length 1464;
 Best Local Similarity 69.2%; Pred. No. 1.1;
 Matches 9; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
 OY 1 GVKGDKGNWGWPG 13
 1111:1:1:1:1
 Db 857 GVKGERGSPGGPG 869

RESULT 30

MSRE_MOUSE

ID MSRE_MOUSE STANDARD; PRT; 458 AA.
 AC P30204; Q9QZ6;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Macrophage scavenger receptor types I and II (Macrophage acetylated
 LDL receptor I and II) (Scavenger receptor type A) (SR-A).

OS M8R1 OR SCVR.
 MUS musculus (Mouse).
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NEBI_TAXID=10090;
 RN [1] SEQUENCE FROM N.A. (ISOFORMS I AND II).
 MEDLINE=93350822; PubMed=8334868;
 RA Ashkenas J., Pennman M., Vasile E., Acton S., Freeman M.W.,
 Krieger M.;
 RT "Structures and high and low affinity ligand binding properties of
 murine type I and type II macrophage scavenger receptors.",
 J. Lipid Res. 34:983-1000(1993).
 RN [2] SEQUENCE FROM N.A. (ISOFORM II).
 RP MEDLINE=93111972; PubMed=6380589;
 RA Doi T., Wada Y., Kodama T., Higashi K.I., Kurihara Y.,
 Miyazaki T., Nakamura H., Uesugi S., Imanishi T., Kawabe Y.,
 Itakura H., Yazaki Y., Matsumoto A.;
 RT "Charged collagen structure mediates the recognition of negatively
 charged macromolecules by macrophage scavenger receptors.",
 J. Biol. Chem. 268:2126-2133(1993).
 RN [3] SEQUENCE FROM N.A. (ISOFORM 1).
 RP STRAIN=C57BL/6;
 RA Rafteri D.L., Whitman S.C., Block A.E., Daugherty A.;
 RT "Identification of a functional domain in class A scavenger receptors
 that mediates metabolism of acLDL.",
 Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
 RN [4] SEQUENCE OF 349-459 FROM N.A.
 RP MEDLINE=9106370; PubMed=1978939;
 RA Freeman M., Ashkenas J., Rees D.J., Kingsley D.M., Copeland N.G.,
 Jenkins N.A., Krieger M.;
 RT "An ancient, highly conserved family of cysteine-rich protein domains
 revealed by cloning type I and type II murine macrophage scavenger
 receptors";
 RT receptors in the macrophage
 PROC. NATL. ACAD. SCI. U.S.A. 87:8810-8814(1990).
 RN [5] SEQUENCE OF 1-4 FROM N.A.
 RP MEDLINE=95395388; PubMed=7166008;
 RA Affring R.P., Freeman M.W.;
 RT "Structure of the murine macrophage scavenger receptor gene and
 evaluation of sequences that regulate expression in the macrophage
 cell line, p18BD.";
 RL Lipid Res. 36:1305-1314(1995).
 CC DEPOSITION OF CHOLESTEROL IN ARTERIAL WALLS DURING ATHEROGENESIS
 TWO TYPES OF RECEPTOR SUBUNITS EXIST. THESE RECEPTORS MEDIATE THE
 ENDOCYTOSIS OF A DIVERSE GROUP OF MACROMOLECULES, INCLUDING
 MODIFIED LOW DENSITY LIPOPROTEINS (LDL).
 CC -I- SUBUNIT: HOMOTRIMER.
 CC -I- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN.
 CC -I- ALTERNATIVE PRODUCTS: 2 ISOFORMS; I (SHOWN HERE) AND II; ARE
 PRODUCED BY ALTERNATIVE SPlicing.
 CC -I- SIMILARITY: CONTAINS 1 SRCR DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 DR EMBL; L04274; AAA39747_1; ALT_INTT.
 DR EMBL; L04275; AAA39748_1; ALT_INTT.
 DR EMBL; D13382; BAA02650_1; --.
 DR EMBL; AF203701; AAF14001_1; --.
 DR EMBL; M59445; AAA37464_1; --.
 DR EMBL; M59446; AAA37465_1; --.
 DR EMBL; U13873; AAC13774_1; --.
 DR PIR; A38260; A38260.
 DR PROSITE; PS00420; SRCR_1; 1.
 DR PROSITE; PS50287; SRCR_2; 1.
 KW Transmembrane; Glycoprotein; Endocytosis; Coiled coil; LDL;
 KW Receptor; Signal-anchor; Alternative splicing.
 FT DOMAIN 1 55 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 56 78 SIGNAL-ANCHOR (TYPE II MEMBRANE PROTEIN)
 FT TRANSMEM (POTENTIAL).
 FT DOMAIN 79 458 (POTENTIAL).
 FT DOMAIN 79 114 EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 209 259 SPACER (PROBABLE).
 FT DOMAIN 277 350 COILED COIL (POTENTIAL).
 FT DOMAIN 357 457 SRCR.
 FT CARBOHYD 94 94 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 107 107 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 147 147 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 188 188 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 253 253 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 271 271 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARSPLC 352 354 TPLP->RSV (IN ISOFORM II).
 FT VARSPLC 355 458 MISSING (IN ISOFORM II).
 FT CONFLICT 47 47 V->L (IN REF. 3).
 FT CONFLICT 110 110 N->K (IN REF. 3).
 FT CONFLICT 120 120 A->E (IN REF. 3).
 FT CONFLICT 130 130 Q->E (IN REF. 3).
 FT CONFLICT 168 168 L->S (IN REF. 3).
 FT CONFLICT 202 202 H->N (IN REF. 3).
 SQ SEQUENCE 458 AA; 50130 MW; 6D4C041C27EE50B CRC64;
 SQ Best Local Similarity 60.3%; Score 55.5; DB 1; Length 458;
 SQ Matches 12; Conservative 50.0%; Pred. No. 0.41; Mismatches 2; Indels 9; Gaps 9;
 SQ Search completed: November 1, 2002, 12:52:53
 Job time : 9 secs

THIS PAGE BLANK (USP TO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model.

Run on: November 1, 2002, 12:19:21 ; Search time 21.5 Seconds

(without alignments) 120.694 Million cell updates/sec

Title: US-09-529-691A-1

Perfect score: 92

Sequence: 1 GVKGDKGNPGWPGAP 15

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : SPTRMBL19;*

- 1: sp_archea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp Rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriaph:*
- 17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	92	100.0	1621	4 Q9h4R9	Q9h4R9 PRELIMINARY; PRT; 1621 AA.
2	69	75.0	886	4 Q9NB7	Q9h4R9; 01-MAR-2001 (TREMBLrel. 16, Created)
3	66	71.7	404	2 Q9FB7	DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
4	64	69.6	673	4 Q1A052	DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
5	64	69.6	1691	11 Q9ESQ2	DE BA472K17.2 (COLLAGEN TYPE IV ALPHA 1) (FRAGMENT).
6	63	68.5	174	13 Q90585	GN COLA1.
7	63	68.5	979	13 Q919K3	OS Homo sapiens (Human).
8	62	67.4	290	5 Q26054	OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
9	62	67.4	412	9 Q9FXY3	OX NEBL_TAXID=9606;
10	62	67.4	1414	5 Q26634	RN [1]
11	62	67.4	1870	4 Q9B0T2	RP SEQUENCE FROM N.A.
12	62	67.4	1802	5 Q1163	RA Bates K.;
13	61	66.3	224	5 Q9VWY2	RL Submitted (OCT-2000) to the EMBL/GenBank/DDBJ databases.
14	61	66.3	447	4 Q16593	DR EMBL; AL390755; CAC13153.1; -
15	61	66.3	957	4 Q9H0V3	DR InterPro; IPR00087; Collagen.
16	66.3	957	4 Q96P44	DR Pfam; PF01413; C4; 2.	
					DR PRODOM; PDD03923; C4; 2.
					DR SMART; SM00111; C4; 2.
					KW Collagen.
					FT NON_TER
					SNQUENCE 1621 AA; 155705 MW; 73RF6FF901CD0EDBA2 CRC64;

QY	1 GVKGDKGNPGWPGAP 15	Query Match Best Local Similarity 100.0%; Pred. No. 1.5e-05; Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
		Db 1215 GVKGDKGNPGWPGAP 1229

RESULT 2

Q9NB7 PRELIMINARY; PRT; 886 AA.

AC	Q9NQB7;	PRELIMINARY;	PRT;	673 AA.
ID	Q14052;			
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)			
DT	01-NOV-1996 (TREMBLrel. 01, Last sequence update)			
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)			
DE	DA2A23.1 (COLLAGEN, TYPE IV, ALPHA 5 (ALPORT SYNDROME))			
DE	(FRAGMENT).			
GN	COL4A5.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarhini; Hominidae; Homo.			
NCBI_TAXID	95656;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Cobley V.,			
RL	Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.			
DR	EMBL: AL03425; CAB90289.1; -.			
DR	InterPro: IPR01442; C4.			
DR	InterPro: IPR00087; Collagen.			
DR	Pfam: PF01413; C4; 2.			
DR	Pfam: PF01391; Collagen; 10.			
DR	ProDom: PD003923; C4; 2.			
DR	SMART: SM00111; C4; 2.			
KW	Collagen.			
FT	NON-TER 1			
SQ	SEQUENCE 886 AA; 85479 MW; 8C06B9FC9AA6569 CRC64;			
Query Match	75.0%; Score 69; DB 4; Length 886;			
Best Local Similarity	73.3%; Pred. No. 0.027;			
Matches	11; Conservative 2; Mismatches 2; Indels 0; Gaps 0;			
OY	1 GVKGDKGNPGWPGAP 15			
Db	480 GIKGEKGKGNPGQGLP 494			
RESULT	3			
ID	09F687	PRELIMINARY;	PRT;	404 AA.
AC	09F687;			
DT	01-MAR-2001 (TREMBLrel. 16, Created)			
DT	01-MAR-2001 (TREMBLrel. 16, Last sequence update)			
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)			
DE	COLLAGEN-LIKE SURFACE PROTEIN (FRAGMENT).			
OS	Streptococcus progenes.			
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;			
OC	Streptococcus.			
OX	NCBI_TAXID=1314;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAM=AP34;			
RX	MLDBN=20590571; PubMed=11035747;			
RA	Rasmussen M., Eden A., Björck L.,			
RT	"ScIA," a novel collagen-like surface protein of streptococcus			
RT	pyogenes, a novel collagen-like surface protein of streptococcus			
RT	infect. Immun. 68:6310-6317(2000).			
RL	EMBL: AF296336; AAG30216.1; -.			
DR	InterPro: IPR00087; Collagen.			
DR	Pfam: PF01391; Collagen; 3.			
FT	NON-TER 1			
FT	NON-TER 404 404			
SQ	SEQUENCE 404 AA; 16162 MW; 45304C89E3B41A78 CRC64;			
Query Match	71.7%; Score 66; DB 2; Length 404;			
Best Local Similarity	73.3%; Pred. No. 0.034;			
Matches	11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;			
OY	1 GVGDKGKGNPGWPGAP 15			
Db	113 GIKGDKGEPGPGLP 127			
RESULT	4			
Query Match	69.6%; Score 64; DB 11; Length 1691;			
Best Local Similarity	66.7%; Pred. No. 0.3;			
Matches	10; Conservative 3; Mismatches 2; Indels 0; Gaps 0;			

OY 1 GVKGDKNPGWPGAP 15
ID |:|:|:||| |
AC 090585 PRELIMINARY; PRT; 174 AA.
DB 1285 GIKGEGKNGPQQQ 1299

RESULT 6

Q90585 PRELIMINARY; PRT; 174 AA.
ID 090585 ID 026054 PRELIMINARY; PRT; 290 AA.
AC 090585; ID 026054; PRELIMINARY; PRT; 290 AA.
DT 01-NOV-1996 (TREMBrel. 01, Created) DT 01-NOV-1996 (TREMBrel. 01, Created)
DT 01-NOV-1996 (TREMBrel. 01, Last sequence update) DT 01-NOV-1996 (TREMBrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBrel. 19, Last annotation update) DT 01-DEC-2001 (TREMBrel. 19, Last annotation update)
DE CARTILAGE ALPHA-1(IX) COLLAGEN PROTEOGLYCAN (FRAGMENT). DE COLLAGEN-LIKE PROTEIN (FRAGMENT).
OS Gallus gallus (Chicken). OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; OC Echinodermata; Echinoidea; Echinozoa;
OC Gallus. OC Paracentrotus lividus (Common sea urchin);
OX NCBI_TAXID:9031; OX Echinidae; Euechinoidea; Echinacea; Echinidae;
RN [1] RN Paracentrotus
RP SEQUENCE FROM N.A. RP NCBI_TAXID:7656;
RX MEDLINE=90052114; PubMed=2584206; RN [1]
RA Nishimura I., Muragaki Y., Olsen B.R.; RP SEQUENCE FROM N.A.
RT "Issue-specific forms of type IX collagen-proteoglycan arise from the use of two widely separated promoters."; RT MEDLINE=89149773; PubMed=2537631;
RL J. Biol. Chem. 264:20033-20041(1989). RA Saitta B., Buttice G., Gambino R.;
DR EMBL: M28660; AAA48705.1; -. RT "Isolation of a putative collagen-like gene from the sea urchin paracentrotus lividus.";
DR InterPro: IPR000087; Collagen. DR Biochem. Biophys. Res. Commun. 158:633-639(1989).
DR Pfam: PF01391; Collagen; 2. DR M24558; AAA2949.1; -;
KW Collagen. DR InterPro: IPR000087; Collagen.
FT NON_TER 174 174 PRT; 174 AA; COE64N0748R224FO CRC64;
SQ SEQUENCE 174 AA; 16240 MW; COE64N0748R224FO CRC64;
Query Match 68.5%; Score 63; DB 13; Length 174; RT NON_TER 1 1
Best Local Similarity 66.7%; Pred. No. 0.041; Length 174; RT NON_TER 1 1
Matches 10; Conservative 3; Mismatches 2; Indels 0; Gaps 0; OX NCBI_TAXID:7656;

Qy 1 GVKGDKNPGWPGAP 15
ID 1:|:|:||| |
AC 46 GIDGDKGSPGAPGPSP 60

RESULT 7

Q9I9K3 PRELIMINARY; PRT; 979 AA.
ID Q9I9K3 ID 026054 PRELIMINARY; PRT; 290 AA.
AC 09I9K3; ID 026054; PRELIMINARY; PRT; 290 AA.
DT 01-OCT-2000 (TREMBrel. 15, Created) DT 01-MAR-2001 (TREMBrel. 16, Created)
DT 01-OCT-2000 (TREMBrel. 15, Last sequence update) DT 01-MAR-2001 (TREMBrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBrel. 17, Last annotation update) DT 01-DEC-2001 (TREMBrel. 19, Last annotation update)
DE COLLAGEN IV A1 CHAIN (FRAGMENT). DE ORF62.
OS Gallus gallus (Chicken). OS bacteriophage phi ETA.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; OC Viruses.
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; OX NCBI_TAXID:9031;
OX NCBI_TAXID:9031; OX NCBI_TAXID:106284;
RN [1] RN [1]
RP SEQUENCE FROM N.A. RP SEQUENCE FROM N.A.
RA Halfter W.M., Dong S.; RA Sugai M., Yamaguchi T., Hayashi T., Nakasone K., Takami H.;
RT "Composition, synthesis and assembly of the embryonic chick retinal basal lamina."; RA Sugai M., Yamaguchi T., Hayashi T., Nakasone K., Ohnishi M.,
RL Dev. Biol. 0:0-0(2000); RA Yamaguchi T., Hayashi T., Takami H., Nakasone K., Ohnishi M.,
DR EMBL: AF239838; AAC44681.1; -. RA Nakayama K., Yamada S., Komatsuwa H., Sugai M.,
DR InterPro: IPR001442; C4. RT Phage conversion of exfoliative toxin A production in Staphylococcus
DR Pfam: PF01413; C4; 2. RT aureus";
DR Pfam: PF01391; Collagen; 12. RL Mol. Microbiol. 38:694-705(2000).
DR Prodrom: PD003923; C4; 2. DR EMBL: AP001553; BAA97648.1; -;
DR SMART: SM00111; C4; 2. DR InterPro: IPR00087; Collagen.
FT NON_TER 1 1 PRT; 979 AA; 95020 MW; 5B1017D911ED4299 CRC64;
SQ SEQUENCE 979 AA; 95020 MW; 5B1017D911ED4299 CRC64;

Query Match 68.5%; Score 63; DB 13; Length 979; RT NON_TER 1 1
Best Local Similarity 60.0%; Pred. No. 0.24; RT NON_TER 1 1
Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0; OX NCBI_TAXID:9031;

RESULT 8

Qy 1 GVKGDKNPGWPGAP 15
ID 1:|:|:||| |
AC 573 GAKGEQSDGWPQTP 587
DB 573 GAKGEQSDGWPQTP 587

Q9FZY3 PRELIMINARY; PRT; 412 AA.
ID Q9FZY3 ID 026054 PRELIMINARY; PRT; 290 AA.
AC 09FZY3; ID 026054; PRELIMINARY; PRT; 290 AA.
DT 01-MAR-2001 (TREMBrel. 16, Created) DT 01-MAR-2001 (TREMBrel. 16, Created)
DT 01-MAR-2001 (TREMBrel. 16, Last sequence update) DT 01-MAR-2001 (TREMBrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBrel. 19, Last annotation update) DT 01-DEC-2001 (TREMBrel. 19, Last annotation update)
DE ORF62. DE ORF62.
OS bacteriophage phi ETA. OS bacteriophage phi ETA.
OC Viruses. OC Viruses.
OX NCBI_TAXID:106284; OX NCBI_TAXID:106284;
RN [1] RN [1]
RP SEQUENCE FROM N.A. RP SEQUENCE FROM N.A.
RA Sugai M., Yamaguchi T., Hayashi T., Nakasone K., Takami H.; RA Sugai M., Yamaguchi T., Hayashi T., Nakasone K., Takami H.;
RL Submitted (MAR-2000) to the EMBL/Genbank/DDBJ databases. RL Submitted (MAR-2000) to the EMBL/Genbank/DDBJ databases.
RN [2] RN [2]
RP SEQUENCE FROM N.A. RP SEQUENCE FROM N.A.
RX MEDLINE=2056787; PubMed=11115106; RX MEDLINE=2056787; PubMed=11115106;
RA Yamaguchi T., Hayashi T., Takami H., Nakasone K., Ohnishi M., RA Yamaguchi T., Hayashi T., Takami H., Nakasone K., Ohnishi M.,
RT Phage conversion of exfoliative toxin A production in Staphylococcus RT Phage conversion of exfoliative toxin A production in Staphylococcus
aureus"; RT Phage conversion of exfoliative toxin A production in Staphylococcus
RL Mol. Microbiol. 38:694-705(2000). RL Mol. Microbiol. 38:694-705(2000).
DR EMBL: AP001553; BAA97648.1; -. DR EMBL: AP001553; BAA97648.1; -;
DR InterPro: IPR00087; Collagen. DR InterPro: IPR00087; Collagen.
SQ SEQUENCE 412 AA; 46013 MW; 3E87FCBAB223088 CRC64;
Query Match 67.4%; Score 62; DB 9; Length 412; RT NON_TER 1 1
Best Local Similarity 78.6%; Pred. No. 0.14; RT NON_TER 1 1
Matches 11; Conservative 0; Mismatches 3; Indels 0; Gaps 0; OX NCBI_TAXID:9031;

Db 178 GAKSDKGEPGQPGA 191
RESULT 10
 ID 026534 PRELIMINARY; PRT; 1414 AA.
 AC Q26534;
 DT 01-NOV-1996 (TREMBREL. 01, Created)
 DT 01-NOV-1996 (TREMBREL. 01, Last sequence update)
 DT 01-DEC-2001 (TREMBREL. 19, Last annotation update)
 DE ALPHA-1 COLLAGEN
 GN COLP1ALPHA.
 OS Strongylocentrotus purpuratus (Purple sea urchin).
 OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
 OC Echinoidea; Euechinoidea; Echinacea; Echinoidea; Strongylocentrotidae;
 OC NCBI_TaxID=7688;
 RN [1]
 RP SEQUENCE FROM N.A.
 MEDLINE=2348411; PubMed=1639795;
 RA Exposito J.-Y.; D'Alessio M.; Solursh M.; Ramirez F.;
 RT "Sea urchin collagen evolutionarily homologous to vertebrate pro-
 alpha-2(1) collagen.";
 J. Biol. Chem. 267:15559-15562 (1992).
 EMBL; M92040; AAA035..1; .
 DR InterPro; IPR000087; Collagen.
 DR InterPro; IPR000885; Fib_collagen_C.
 PFam; PF01410; COLF1; 1.
 PFam; PF01391; Collagen; 18.
 DR probom; PD002078; Fib_collagen_C; 1.
 DR SMART; SM00038; COLF1; 1.
 DR Collagen.
 KW SEQUENCE 1414 AA; 133025 MW; 94D9CBA71A9FD73D CRC64;
 Query Match 67.4%; Score 62; DB 5; Length 1414;
 Best Local Similarity 78.6%; Pred. No. 0.51;
 Matches 11; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 Qy 1 GVKDQKGNPGWPGA 14
 | |||:||||| |||||
 Db 393 GSKEGDQGNPGQPGA 406
 OS Homo sapiens (Human).
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21064696; PubMed=11134255;
 RA Heidet L., Arredel C., Forestier L., Cohen-Solal L., Mollet G.,
 Gutierrez B., Stavrou C., Guiller M.C., Antignac C.;
 RT "Structure of the human type IV collagen gene COL4A3 and mutations in
 autosomal Airport syndrome.";
 J. Am. Soc. Nephrol. 12:97-106 (2001).
 EMBL; AJ288487; CAC36101..1; .
 DR AJ288488; CAC36101..1; JOINED.
 EMBL; AJ288489; CAC36101..1; JOINED.
 DR AJ288490; CAC36101..1; JOINED.
 EMBL; AJ288491; CAC36101..1; JOINED.
 DR AJ288492; CAC36101..1; JOINED.
 EMBL; AJ288493; CAC36101..1; JOINED.
 EMBL; AJ288494; CAC36101..1; JOINED.

RESULT 11
 ID Q9BQ72 PRELIMINARY; PRT; 1670 AA.
 AC Q9BQ72;
 DT 01-JUN-2001 (TREMBREL. 17, Created)
 DT 01-OCT-2001 (TREMBREL. 17, Last sequence update)
 DT 01-OCT-2001 (TREMBREL. 18, Last annotation update)
 DE ALPHA3 TYPE IV COLLAGEN.
 COLA43
 OS Homo sapiens (Human).
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21064696; PubMed=11134255;
 RA Heidet L., Arredel C., Forestier L., Cohen-Solal L., Mollet G.,
 Gutierrez B., Stavrou C., Guiller M.C., Antignac C.;
 RT "Structure of the human type IV collagen gene COL4A3 and mutations in
 autosomal Airport syndrome.";
 J. Am. Soc. Nephrol. 12:97-106 (2001).
 EMBL; AJ288487; CAC36101..1; .
 DR AJ288488; CAC36101..1; JOINED.
 EMBL; AJ288489; CAC36101..1; JOINED.
 DR AJ288490; CAC36101..1; JOINED.
 EMBL; AJ288491; CAC36101..1; JOINED.
 DR AJ288492; CAC36101..1; JOINED.
 EMBL; AJ288493; CAC36101..1; JOINED.
 EMBL; AJ288494; CAC36101..1; JOINED.

RESULT 12
 ID 017163 PRELIMINARY; PRT; 1802 AA.
 AC 017163;
 DT 01-NOV-1996 (TREMBREL. 01, Created)
 DT 01-NOV-1996 (TREMBREL. 01, Last sequence update)
 DT 01-DEC-2001 (TREMBREL. 19, Last annotation update)

DE A2 (IV) BASEMENT MEMBRANE COLLAGEN.	RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshefi A.,
OS Brugia malayi.	RA Mount S.M., Moy M., Murphy B., Murphy L., Muzyk D.M., Nelson D.L.,
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;	RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
OC Onchocercidae; Brugia.	RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
OX NCBI_TAXID=6279;	RA Reinert K., Remington K., Saunders R.D.C., Schaefer F., Shen H.,
[1]	RA Shue B.C., Sider-Klamos I., Simpson M., Skupski M.P., Smith T.,
RP SEQUENCE FROM N.A.	RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RC STRAIN=APERTOIDIC;	RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RX MEDLINE=9534846; Published=7637709;	RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
RA Caulagi V.R., Rajan T.V.,	RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RT "The structural organization of an alpha 2 (type IV) basement membrane	RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao O., Zheng L.,
collagen gene from the filarial nematode <i>Brugia malayi</i> .";	RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RL Mol. Biochem. Parasitol. 70:227-229(1995).	RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
DR EMBL; U07224; AAC6611.1; -.	RA RT "The genome sequence of <i>Drosophila melanogaster</i> .";
DR InterPro; IPR001442; C4.	RA Science 287:2185-2195(2000).
DR InterPro; IPR00087; Collagen.	RA EMBL; AB003517; ARF49173.1.
DR Pfam; PF01413; C4; 2.	RA DR FlyBase; FBgn0056861; CG14089.
DR ProDom; PDO03923; C4; 2.	RA DR InterPro; IPR00087; Collagen.
DR SMART; SM00111; C4; 2.	RA DR InterPro; IPR02665; PRich_Extensn.
KW Collagen.	RA DR PRINTS; PR01391; Collagen; 1.
SEQUENCE 1802 AA; 172401 MW; 595P16554CBE2D24 CRC64;	RA DR SQ SEQUENCE 224 AA; 24308 MW; F84B9C912D8EC1CD CRC64;
Query Match 67.4%; Score 62; DB 5; Length 1802;	Query Match 66.3%; Score 61; DB 5; Length 224;
Best Local Similarity 71.4%; Pred. No. 0.65; 3; Mismatches	Best Local Similarity 66.7%; Pred. No. 0.11; 2; Mismatches
Matches 10; Conservative 3; Indels 1; Gaps 0; Gaps 0;	Matches 11; Conservative 3; Indels 0; Gaps 0; Gaps 0;
Qy 1 GVKGDKGNGPGWCA 14	Qy 1 GVKGDKGNGPGWCA 15
Db 1196 GRGDKGSPGPGCA 1209	Db 118 GRGDKGSPGPGCA 132
RESULT 13	RESULT 14
09vvy2 PRELIMINARY; PRT; 224 AA.	Q16593 PRELIMINARY; PRT; 447 AA.
ID 09vvy2	ID 016593
AC	AC 016593:
DT 01-MAY-2000 (TREMBREL 13, Created)	DT 01-NOV-1996 (TREMBREL 01, Created)
DT 01-MAY-2000 (TREMBREL 13, Last sequence update)	DT 01-NOV-1996 (TREMBREL 01, Last sequence update)
DT 01-JUN-2001 (TREMBREL 17, Last annotation update)	DT 01-JUN-2001 (TREMBREL 17, Last annotation update)
DE CG14089 PROTEIN.	DE DE COLLAGEN-LIKE PROTEIN (47 AA) (FRAGMENT).
GN CG14089.	OS Homo sapiens (Human).
OS Drosophila melanogaster (FRUIT FLY).	OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;	OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC Pierigota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;	OX NEBI_TAXID=9606;
OC Ephydriidae; Drosophilidae; Drosophila.	RN [1]
OX NEBI_TAXID=227;	RP SEQUENCE FROM N.A.
RN [1]	RC TISSUE=PLACENTA;
RP SEQUENCE FROM N.A.	RA Kimura S.;
RC STRAIN=BERKELEY;	RL Submitted (FEB-1990) to the EMBL/GenBank/DDBJ databases.
RX MEDLINE=20196006; Published=10731132;	RN [2]
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,	RP SEQUENCE FROM N.A.
RA Ananatis P.G., Scheer S.E., Li P.W., Hoskins R.A., Galle R.F.,	RC TISSUE=PLACENTA;
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,	RA Kimura S.;
RA Sutton G.G., Wormit J.R., Yandell M.D., Zhang Q., Chen L.X.,	RL Submitted (MAY-1992) to the EMBL/GenBank/DDBJ databases.
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Chapple M., Preiffer B.D.,	DR EMBL; X14963; CA33085.1; -.
RA Wan K.H., Doyle C., Baxter E., Heit G., Nelson C.R., Miklos G.L.G.,	DR EMBL; X15038; CA33142.1; -.
RA Abril J.F., Agbayani A., An H.-J., Andrewn-Pfaunkoch C., Baldwin D.,	DR InterPro; IPR00087; Collagen.
RA Balieu R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,	DR Pfam; PF01391; Collagen; 5.
RA Beeson K.Y., Benos P.V., Bernick B.P., Bhandari D., Bolshawik S.,	FT NON_TER 1
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottler P.,	FT NON_TER 447 447
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra T.,	SQ SEQUENCE 447 AA; 41829 MW; FDB207023db7CC94 CRC64;
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,	Query Match 66.3%; Score 61; DB 4; Length 447;
RA de Pablo B., Deicher A., Delleck Z., Mays A.D., Dew I., Dietz S.M.,	Best Local Similarity 73.3%; Pred. No. 0.22; 4; Indels 0; Gaps 0;
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,	Matches 11; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,	Qy 1 GVKGDKGNGPGWCA 15
RA Fosler C., Gabrieleian A.E., Garg N.S., Geibart W.M., Glasser K.,	Db 280 GOKGDAGNPQPGTP 294
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.J.,	RESULT 15
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,	RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,	RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A., Li J., Li Z., Liang Y., Lin X.,	RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,	RA Q9H0V3

ID	Q9H0V3	PRELIMINARY;	PRT;	957 AA.		
AC	Q9H0V3;0					
DT	01-MAR-2001	(TREMBREL. 16, Created)				
DT	01-MAR-2001	(TREMBREL. 16, Last sequence update)				
DE	HYPOTHETICAL	99.4 KDA PROTEIN.				
GN	DKF2P5641052.					
OS	Homo sapiens (Human).					
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
OX	NCBI_TAXID=9606;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RC	TISSUE=BRAIN;					
RX	MEDLINE=21154917; PubMed=11230166;					
RA	Wiemann S., Well B., Weilenreuther R., Gassnerhuber J., Glassl S.,					
RA	Ansorge W., Beecher M., Bloecker H., Bauersachs S., Blum H.,					
RA	Laufer J., Duesterhoeft A., Beyer A., Koehler K., Strack N.,					
RA	Mewes H.W., Oettwelder B., Obermaier B., Tampe J., Heubner D.,					
RA	Wambutt R., Korn B., Klein M., Pousta K.; Tampe J., Heubner D.,					
RT	"Towards a Catalog of Human Genes and Proteins: Sequencing and Analysis of 500 Novel Complete Protein Coding Human cDNAs".					
RL	Genome Res. 11:42-43(2001).					
DR	EMBL; ALI36624; CAB66559.1; -.					
DR	HSSP; P17301; IAOX.					
DR	InterPro; IPR00087; Collagen.					
DR	InterPro; IPR003129; TSN.					
DR	InterPro; IPR002035; vWFA.					
DR	Pfam; PF01391; Collagen; 6.					
DR	Pfam; PF02210; TSPN; 1.					
DR	Pfam; PF00092; vWA; 1.					
DR	PRINTS; PR00453; vWFADOMAIN.					
DR	SMART; SM00210; TSPN; 1.					
DR	SMART; SM03271; vWFA; 1.					
DR	PROSITE; PS50234; vWFA; 1.					
KW	Hypothetical protein.					
SQ	SEQUENCE 957 AA; 99413 MW; 391022F715024571 CRC64;					
QY	1 GVKGDKGNPGWPGAP 15					
DB	551 GAKSEKGNAGFPGLP 565					
Query Match Similarity 66.3%; Score 61; DB 4; Length 957; Best Local Similarity 66.7%; Pred. No. 0.48; Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;						
RESULT 16						
ID	096P44	PRELIMINARY;	PRT;	957 AA.		
AC	Q96P44;0					
DT	01-DEC-2001	(TREMBREL. 19, Created)				
DT	01-DEC-2001	(TREMBREL. 19, Last sequence update)				
DT	01-DEC-2001	(TREMBREL. 19, Last annotation update)				
DE	COLLAGEN XXI.					
GN	COL2A1.					
OS	Homo sapiens (Human).					
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
OX	NCBI_TAXID=9606;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RA	Brown A.;					
RL	Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.					
DR	EMBL; ALI38761; CAB00589.1; -.					
DR	InterPro; IPR00087; Collagen.					
DR	Pfam; PF01391; Collagen; 7.					
DR	PF01391; Collagen.					
SQ	SEQUENCE 1497 AA; 150419 MW; E01027005F3AE843 CRC64;					
QY	1 GVKGDKGNPGWPGA 14					
DB	960 GPKDGDGDPGVGA 973					
Query Match Similarity 65.2%; Score 60; DB 4; Length 1497; Best Local Similarity 78.6%; Pred. No. 1.1; Matches 11; Conservative 1; Mismatches 2; Indels 0; Gaps 0;						
QY	1 GVKGDKGNPGWPGA 15					

		RESULT 19	Db	960	GPKGDKGDFGVPGA	973
Q30MD9	ID	PRELIMINARY;	PRN;	1497	AA.	
Q30MD9	ID	PRELIMINARY;	PRN;	1532	AA.	RESULT 20
Q30MD9	ID	PRELIMINARY;	PRN;	1532	AA.	
DT	01-MAY-2000	(TREMBlrel. 13, Created)	AC	Q02802		
DT	01-MAY-2000	(TREMBlrel. 17, Last annotation update)	AC	Q02802;		
DE	180	KDA BULLOUS PEMPHIGOID ANTIGEN 2/TYPE XVII COLLAGEN.	AC	Q02802;		
GN	BPG2/COL17A1		DT	01-NOV-1996	(TREMBlrel. 01, Last sequence update)	
OS	Homo sapiens (Human).		DT	01-NOV-1996	(TREMBlrel. 01, Last sequence update)	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		DT	01-DEC-2001	(TREMBlrel. 19, Last annotation update)	
OC	Mammalia; Eutheria; Primates; Catarrini; Homidae; Homo.		DE	AUTOTANTIGEN (FRAGMENT).		
OX	NBBI_TAXID=9606;		GN	BP180.		
RN	[1]		OS	Homo sapiens (Human).		
RP	SEQUENCE FROM N.A.		OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
RX	MEDLINE=97164601; PubMed=9012408;		OC	Mammalia; Eutheria; Primates; Catarrini; Homidae; Homo.		
RA	Gatalica B., Pulkkinen L., Li K., Kuokkanen K., Rynnanen M., McGrath J.A., Uitto J.:		OX	NBBI_TAXID=9606;		
RT	"Cloning of the human type XVII collagen gene (COL17A1), and detection of novel mutations in generalized atrophic benign epidermolysis bullosa".		RN	[1]		
RL	Am. J. Hum. Genet. 60:352-365(1997).		RP	SEQUENCE FROM N.A.		
DR	EMBL; 076604; AA51499_1; JOINED.		RC	TISSUE-FORESKIN;		
DR	EMBL; 076565; AA51499_1; JOINED.		RX	MEDLINE=92381323; PubMed=1324962;		
DR	EMBL; 076567; AA51499_1; JOINED.		RA	Gludice G.J.; Emery D.J.; Diaz L.A.;		
DR	EMBL; 076568; AA51499_1; JOINED.		RT	"Cloning and Primary structural analysis of the Bullous pemphigoid autoantigen. BP180.";		
DR	EMBL; 076569; AA51499_1; JOINED.		RT	autoantigen. BP180.;		
DR	EMBL; 076570; AA51499_1; JOINED.		DR	J. Invest. Dermatol. 99:243-250(1992).		
DR	EMBL; 076571; AA51499_1; JOINED.		DR	EMBL; M01668; AAA35605_1; .		
DR	EMBL; 076572; AA51499_1; JOINED.		DR	InterPro; IPR000087; Collagen.		
DR	EMBL; 076573; AA51499_1; JOINED.		FT	Prfam; PF01391; Collagen; 6.		
DR	EMBL; 076574; AA51499_1; JOINED.		FT	NON-TER	1	
DR	EMBL; 076575; AA51499_1; JOINED.		SEQUENCE	1532 AA;	154568 MW;	BF35054CF93B8EB3 CRC64;
DR	EMBL; 076576; AA51499_1; JOINED.		Db	995	GPKGDKGDFGVPGA	1008
DR	EMBL; 076577; AA51499_1; JOINED.		Query Match	65.2%	Score 60;	DB 4;
DR	EMBL; 076578; AA51499_1; JOINED.		Best Local Similarity	78.6%	Pred. No. 1.1;	Length 1532;
DR	EMBL; 076579; AA51499_1; JOINED.		Matches	11;	Mismatches	1;
DR	EMBL; 076580; AA51499_1; JOINED.		Qy	1	GVKGDKGNFGWPGA	14
DR	EMBL; 076581; AA51499_1; JOINED.			^ ^ ^ ^ ^ : ^ ^ ^ ^ ^		
DR	EMBL; 076582; AA51499_1; JOINED.					
DR	EMBL; 076583; AA51499_1; JOINED.					
DR	EMBL; 076584; AA51499_1; JOINED.					
DR	EMBL; 076585; AA51499_1; JOINED.					
DR	EMBL; 076586; AA51499_1; JOINED.					
DR	EMBL; 076587; AA51499_1; JOINED.					
DR	EMBL; 076588; AA51499_1; JOINED.					
DR	EMBL; 076589; AA51499_1; JOINED.					
DR	EMBL; 076590; AA51499_1; JOINED.					
DR	EMBL; 076591; AA51499_1; JOINED.					
DR	EMBL; 076592; AA51499_1; JOINED.					
DR	EMBL; 076593; AA51499_1; JOINED.					
DR	EMBL; 076594; AA51499_1; JOINED.					
DR	EMBL; 076595; AA51499_1; JOINED.					
DR	EMBL; 076596; AA51499_1; JOINED.					
DR	EMBL; 076597; AA51499_1; JOINED.					
DR	EMBL; 076598; AA51499_1; JOINED.					
DR	EMBL; 076599; AA51499_1; JOINED.					
DR	EMBL; 076600; AA51499_1; JOINED.					
DR	EMBL; 076601; AA51499_1; JOINED.					
DR	EMBL; 076602; AA51499_1; JOINED.					
DR	EMBL; 076603; AA51499_1; JOINED.					
DR	InterPro; IPR000087; Collagen.					
KW	Pfam; PF01391; Collagen; 6.					
KW	Collagen.					
SQ	SEQUENCE FROM N.A.					
Query Match	65.2%	Score 60;	DB 4;	Length 1497;		
Best Local Similarity	78.6%	Pred. No. 1.1;				
Matches	11;	Conservative	1;			
Indels	2;	Indels	0;			
Gaps	0;	Gaps	0;			
Gaps	0;	Gaps	0;			
Qy	1	GVKGDKGNFGWPGA	14			
Qy	1	GVKGDKGNFGWPGA	15			

Db	994 GRKGEGTGNPGFPGRP 1008	Qy	1 GVKGDKGNPGMPGAP 15
RESULT 22	Q9VMV5	PRELIMINARY;	PRT; 1940 AA.
ID	Q9VMV5;		
AC	01-MAY-2000 (T-EMBLrel. 13, Created)	ID	063870
DT	01-MAY-2000 (T-EMBLrel. 13, Last sequence update)	AC	063870;
DT	01-JUN-2001 (T-EMBLrel. 17, Last annotation update)	DT	01-NOV-1998 (T-EMBLrel. 08, Last sequence update)
DE	VKG PROTEIN.	DT	01-DEC-2001 (T-EMBLrel. 19, Last annotation update)
GN	VKG OR CG16858.	DE	TYPE VII COLLAGEN.
OS	drosophila melanogaster (Fruit fly).	GN	TYPE VII COLLAGEN.
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;	OS	Mus musculus (Mouse).
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;	OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OX	Ephydriidae; Diopsophilidae; Drosophila.	OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murine; Mus.
RN	[1]	OX	NCBI_TaxID=1008;
RP	SEQUENCE FROM N.A.	RN	[1]
RC	STRAIN=BERKELEY;	RP	SEQUENCE FROM N.A.
RX	MEDLINE=20006; PubMed=10731132;	RX	MEDLINE=9331518; PubMed=8325648;
RA	Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D., Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F., George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N., Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X., Brandt R.C., Rogers Y.H.C., Blazier R.G., Champe M., Pfeiffer B.D., Wan K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Miklos G.L.G., Abril J.F., Agayam A., An H.-J., Andrews-Pfannkoch C., Baldwin D., Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M., Beeson K.Y., Benos P.V., Bernick B.P., Bhandari D., Bolshakov S., Borckova D., Bozchian M.R., Bouck J., Brokstein P., Bröttner P., Burtis K.C., Busam D.A., Butler H., Cadieu E., Chandra I., Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P., de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M., Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P., Durbin K.J., Evangelista C.C., Ferraz C., Fleischmann W., Fosler C., Gabrieloff A.E., Garg N.S., Gelbart W.M., Glasser K., Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M., Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J., Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C., Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A., Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulpa D., Lal Z., Lasko P., Lei Y., Levitsky D., Li J., Li Z., Liang Y., Lin X., Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D., Merkulov G., Milashina N.V., Mobarry C., Morris J., Moskrafi A., Mount S.M., Moy M., Murphy B., Murphy L., Muzyk D.M., Nelson D.L., Nelson D.R., Nelson K.A., Nusskern D.R., Paclet J.M., Palazolo M., Pittman G.S., Pan S., Pollard J., Purifoy E., Reese M.G., Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H., Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T., Spier E., Spradling A.C., Stapleton M., Strong R., Sun E., Svirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X., Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissenbach J., Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A., Yeh J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao O., Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.; "The genome sequence of Drosophila melanogaster.", Science 287:2185-2195(2000). EMBL: AE003608; AAC02203.1; -. DR FLYBase; FBgn0016075; vkg. DR Interpro; IPR001442; C4. DR Interpro; IPR000087; Collagen. DR Pfam; PF01413; C4; 2. DR Pfam; PF01391; Collagen; 21. DR SMART; SW00393; C4; 2. DR SMART; SW00111; C4; 2. DR SEQUENCE 1940 AA; 193777 MW; 9B507382EF9C17B5 CRC64;	Qy	1 GVKGDKGNPGMPGAP 15
SQ	Query Match 65.2%; Best Local Similarity 73.3%; Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;	DR	2377 GMKGDLGPRAPGAP 2391
Query Match 65.2%; Score 60; DB 5; Length 1940; Best Local Similarity 66.7%; Pred. No. 1.4; Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;	RESULT 24	026312	PRELIMINARY; PRT; 907 AA.

AC Q26312; DT 01-NOV-1996 (TREMBREL_01, Created) ID Q9UJC7 PRELIMINARY; PRT; 771 AA. DE DT 01-NOV-1996 (TREMBREL_01, Last sequence update) ID Q9UJC7 PRELIMINARY; PRT; 771 AA. DE NONFIBRILLAR COLLAGEN PROTEIN (FRAGMENT). DT 01-MAY-2000 (TREMBREL_13, Created) ID Q9UJC7 PRELIMINARY; PRT; 771 AA. DE OS Strongylocentrotus purpuratus (Purple sea urchin). DT 01-MAY-2000 (TREMBREL_13, Last sequence update) ID Q9UJC7 PRELIMINARY; PRT; 771 AA. DE OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa; OC Echinida; Euechinida; Echinoidea; Echinacea; Echinolla; Strongylocentrotidae; OC Strongylocentrotus. DB NCBITAXID=7668; [1] Sequence from N.A.

RP MEDLINE=9203839; PubMed=1936564; RX Wessel G.M.; Etkin M.; Benson S.; RA "Primary mesenchyme cells of the sea urchin embryo require an RT autonomously produced, nonfibrillar collagen for spiculogenesis."; RL Dev. Biol. 148:261-272(1991). DR EMBL: S64572; AAC020270; -; DR InterPro: IPR000087; Collagen. PRfam: PF01391; Collagen; 13. KW Collagen. FT NON_TER 1 SQ SEQUENCE 907 AA; '85400 MW; B8ECBF6FDDA59D88 CRC64; OQ 1 GVKGDDPGRTGPEGAKGNGLPGIP 15 Db 546 GVKGDPGPRTGPEGAKGNGLPGIP 569

RESULT 25

ID 007265 PRELIMINARY; PRT; 1752 AA. AC 007265; DT 01-NOV-1996 (TREMBREL_01, Created) ID Q9UJC7 PRELIMINARY; PRT; 771 AA. DE DT 01-NOV-1996 (TREMBREL_01, Last sequence update) ID Q9UJC7 PRELIMINARY; PRT; 771 AA. DE 3 ALPHA PROCOLLAGEN. COL3ALPHA. OS Strongylocentrotus purpuratus (Purple sea urchin). Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa; OC Echinida; Euechinida; Echinoidea; Echinacea; Echinolla; Strongylocentrotidae; OC Strongylocentrotus. NCBI_TAXID=7668; RN [1]

RP Sequence from N.A. MEDLINE=93186842; PubMed=8444899; RA Exposito J.-Y.; Di Liberto M.; Ramirez F.; RT "Complete primary structure of a sea-urchin type IV collagen and analysis of the 5' end of its gene."; RL Biol. Chem. 268:5249-5254(1993). DR EMBL: L02917; AAC30039; 1; -; DR InterPro: IPR001442; C4. DR Pfam: PF01413; C4; 2. DR Pfam: PF01391; Collagen; 22. DR Pfam: PF003923; C4; 2. DR SMART: SM00111; C4; 2. DR KW Collagen. SEQUENCE FROM N.A. 1752 AA; 170210 MW; 1AE5AAA21569346D CRC64;

Query Match 64.7%; Score 59.5; DB 5; Length 1752; Best Local Similarity 54.2%; Pred. No. 1.5; Matches 13; Conservative 0; Mismatches 2; Indels 9; Gaps 1; PRT; 1433 AA.

OQ 1 GVKGDDPGRTGPEGAKGNGLPGIP 15 Db 1177 GVKGDPGPRTGPEGAKGNGLPGIP 1200

RESULT 26

ID Q9UJC7 PRELIMINARY; PRT; 771 AA. AC 007563 PRELIMINARY; PRT; 1433 AA. DE DT 01-JUN-2001 (TREMBREL_17, Last sequence update) ID Q9UJC7 PRELIMINARY; PRT; 771 AA. DE COI4A5 (COLLAGEN ALPHA 5(IV) CHAIN PRECURSOR) (FRAGMENT). GN COI4A5. OS Homo sapiens (Human). OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrini; Homiidae; Homo. OC NCBI_TAXID=9606; RN [1]

RP Sequence from N.A. RA Bird C.; DR Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases. RL EMBL: AL031622; CAA20937; 1; DR InterPro: IPR000087; Collagen. PRfam: PF01391; Collagen; 10. KW Collagen. FT NON_TER 1 SQ SEQUENCE 771 AA; 72760 MW; 9CFCD9373E545A61 CRC64; OQ 1 GVKGDKGNPGWPGCAP 15 Db 528 GMKGDKGELGSPGAP 542

RESULT 27

ID Q9UJC4 PRELIMINARY; PRT; 1431 AA. AC 007564; DT 01-OCT-2000 (TREMBREL_15, Created) ID Q9UJC4 PRELIMINARY; PRT; 1431 AA. DE DT 01-OCT-2000 (TREMBREL_15, Last sequence update) ID Q9UJC4 PRELIMINARY; PRT; 1431 AA. DE DT 01-JUN-2001 (TREMBREL_17, Last annotation update) ID Q9UJC4 PRELIMINARY; PRT; 1431 AA. DE COLLAGEN TYPE XVII. OS Mesocricetus auratus (Golden hamster). Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae; OC Mesocricetus. NCBI_TAXID=10036; RN [1]

RP Sequence from N.A. Yamamoto K.; Inoue N.; Fujimori A.; Saito T.; Shinkai H.; Sakiyama H.; RA "Mesocrietus auratus mRNA for type XVII collagen."; RT Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases. RL EMBL: AB027159; BAB49430; 1; DR InterPro: IPR000087; Collagen. DR Pfam: PF01391; Collagen; 5. DR SEQUENCE 1431 AA; 144579 MW; 4315631FEB2C9ASC CRC64; OQ 1 GVKGDKGNPGWPGCAP 15 Db 945 GPKGDPGPRTGPEGAKGNGLPGIP 959

RESULT 28

ID Q9UJC7 PRELIMINARY; PRT; 1433 AA. AC 007563 PRELIMINARY; PRT; 1433 AA. DE DT 01-NOV-1996 (TREMBREL_01, Created) ID Q9UJC7 PRELIMINARY; PRT; 1433 AA. DE DT 01-JUN-2001 (TREMBREL_17, Last sequence update)

DE PROCOLLAGEN, TYPE XVII, ALPHA 1 (BULLOUS PEMPHIGOID AUTOANTIGEN
BP180) (COLLAGEN XVII).
DE COL17AI.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TAXID=10090;
RN [1]
SEQUENCE FROM N.A.
STRAIN=BALB/C;
RX MEDLINE=93232041; PubMed=8473327;
RA Li K., Tamai K., Iain E.M., Utton J.;
RT "Cloning of type XVII collagen. Complementary and genomic DNA
sequences of mouse 180-kilodalton bullous pemphigoid antigen (BPAG2)
predict an interrupted collagenous domain, a transmembrane segment,
and unusual features in the 5'-end of the gene and the 3'-untranslated
region of the mRNA.";
RL J. Biol. Chem. 268:8825-8834(1993).
RT -!- FUNCTION: THE COLLAGEN DOMAINS OF BP180 ALLOW THE HEMIDESMOSOME TO
FORM STABLE INTERACTIONS WITH THE CONSTITUENTS OF THE EXTRACELLULAR
MATRIX OF THE CUTANEOUS BASEMENT MEMBRANE ZONE.
CC -!- SUBCELLULAR LOCATION: A COMPONENT OF THE HEMIDESMOSOME.
CC -!- DISEASE: UPON DISRUPTION OF BP180 BY AUTOANTIBODIES THE
HEMIDESMOSOME IS DISRUPTED LEADING TO THE BLISTERING SKIN DISORDER
CC BULLOUS PEMPHIGOID (BP).
DR EMBL; LO4407; AAA37435; 1; -.
DR MGD; MGI:188450; Col17ai.
DR InterPro; IPR000087; Collagen.
DR Pfam; PF01391; Collagen; 5.
KW Antigen; Cell adhesion.
SEQUENCE 1433 AA: 144087 MW: B8F808832A19922C CRC64;
Query Match 64.1%; Score 59; DB 11; Length 1433;
Best Local Similarity 66.7%; Pred. No. 1.5;
Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
QY 1 GVKGDKGNPGWPG 15
| ||||:||| | | |
Db 948 GPKDQGDPGVPGP 962

RESULT 30
Q9J104 PRELIMINARY;
ID Q9J104 PRT; 1737 AA.
AC Q9J104;
DT 01-OCT-2000 (TREMBREL. 15, Created)
DT 01-OCT-2000 (TREMBREL. 15, Last sequence update)
DT 01-DEC-2001 (TREMBREL. 19, Last annotation update)
DE ALPHA 4 TYPE V COLLAGEN
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TAXID=10116;
RN [1]
SEQUENCE FROM N.A.
STRAIN=SPRAGUE DAWLEY;
RX MEDLINE=20428740; PubMed=10852920;
RA Cherenkov M.A., Rothblum K., Tyler W.A., Stahl R.C., Carey D.J.; alpha
RT "Schwann cells synthesize type V collagen that contains a novel alpha
4 chain. molecular cloning, biochemical characterization, and high
affinity heparin binding of alpha(V) collagen.";
RT J. Biol. Chem. 275:28208-28215(2000).
DR EMBL; AF272661; AAC76432; 1; -.
DR InterPro; IPR000087; Collagen.
DR InterPro; IPR000885; Fib.collagen_C.
DR InterPro; IPR001791; Laminin_G.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00383; ColfI; 1.
DR SMART; SM00282; Lang; 1.
DR SMART; SM00210; TSPN; 1.
KW Collagen.
SQ SEQUENCE 1737 AA; 171574 MW; D635D5D57481C257 CRC64;

Q9Q2R9 PRELIMINARY;
ID Q9Q2R9; PRT; 1682 AA.
AC Q9Q2R9;
DT 01-MAY-2000 (TREMBREL. 13, Created)
DT 01-JUN-2001 (TREMBREL. 17, Last annotation update)
DE ALPHA 4 COLLAGEN IV.
GN COL4A4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBL_TAXID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-KIDNEY;
RX MEDLINE=20005934; PubMed=10534397;
RA Lu W., Phillips C.L., Killen P.D., Hlaing T., Harrison W.R.,
RA Elder F.F.B., Miner J.H., Overbeek P.A., Meissler M.H.;
RT "Intragenic mutation of the collagen genes col4a3 and col4a4 in a
mouse model of alport syndrome.";
RL Genomics 61:113-124(1999).
DR EMBL; AF169388; AAC05450; 1; -.
MGD; MGI:104687; Col4a4.
DR InterPro; IPR001442; Col4a4.
DR InterPro; IPR000087; Collagen.
DR Pfam; PF01413; Col4a4.
DR Pfam; PF01391; Collagen; 21.
DR ProDom; PD003923; Col4a4.
DR SMART; SM00111; Col4a4.
KW Collagen. 1682 AA; 164096 MW; 6F7B679EDD76E904 CRC64;

Query Match 64.1%; Score 59; DB 11; Length 1682;
Best Local Similarity 60.0%; Pred. No. 1.7;
Matches 9; Conservative 4; Mismatches 2;
Indels 0; Gaps 0;

QY 1 GVKGDKGNPGWPG 15
| :|||:||| | | |
Db 176 GIQQDRGDPGPPGP 190

Query Match 64.1%; Score 59; DB 11; Length 1737;
Best Local Similarity 76.9%; Pred. No. 1.8;
Matches 10; Conservative 1; Mismatches 2;
Indels 0; Gaps 0;

QY 1 GVKGDKGNPGWPG 13
| ||||:||| | | |
Db 740 GPKDQRGNPGLP 752

Search completed: November 1, 2002, 12:53:44
Job time : 24.5 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

On protein - protein search, using SW model.
Run on: November 1, 2002, 12:14:41. ; Search time 26.5 seconds

(without alignments) updates/sec
62.872 Million cell

Title: US-09-529-691A-3
Perfect score: 92
Sequence: 1 PAGPGPGNGKDGKVG 15
Scoring table: BL0SUM62
Gapop 10.0 , Gapext 0.5
Searched: 74574 seqs, 11073796 residues
Total number of hits satisfying chosen parameters: 74574
Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : A_Geneseq_032802:
1: /SIDS1/gcadata/geneseq/geneseq/emb1/AA1980.DAT: *
2: /SIDS1/gcadata/geneseq/geneseq/emb1/AA1981.DAT: *
3: /SIDS1/gcadata/geneseq/geneseq/emb1/AA1982.DAT: *
4: /SIDS1/gcadata/geneseq/geneseq/emb1/AA1983.DAT: *
5: /SIDS1/gcadata/geneseq/geneseq/emb1/AA1984.DAT: *
6: /SIDS1/gcadata/geneseq/geneseq/emb1/AA1985.DAT: *
7: /SIDS1/gcadata/geneseq/geneseq/emb1/AA1986.DAT: *
8: /SIDS1/gcadata/geneseq/geneseq/emb1/AA1987.DAT: *
9: /SIDS1/gcadata/geneseq/geneseq/emb1/AA1988.DAT: *
10: /SIDS1/gcadata/geneseq/geneseq/emb1/AA1989.DAT: *
11: /SIDS1/gcadata/geneseq/geneseq/emb1/AA1990.DAT: *
12: /SIDS1/gcadata/geneseq/geneseq/emb1/AA1991.DAT: *
13: /SIDS1/gcadata/geneseq/geneseq/emb1/AA1992.DAT: *
14: /SIDS1/gcadata/geneseq/geneseq/emb1/AA1993.DAT: *
15: /SIDS1/gcadata/geneseq/geneseq/emb1/AA1994.DAT: *
16: /SIDS1/gcadata/geneseq/geneseq/emb1/AA1995.DAT: *
17: /SIDS1/gcadata/geneseq/geneseq/emb1/AA1996.DAT: *
18: /SIDS1/gcadata/geneseq/geneseq/emb1/AA1997.DAT: *
19: /SIDS1/gcadata/geneseq/geneseq/emb1/AA1998.DAT: *
20: /SIDS1/gcadata/geneseq/geneseq/emb1/AA1999.DAT: *
21: /SIDS1/gcadata/geneseq/geneseq/emb1/AA2000.DAT: *
22: /SIDS1/gcadata/geneseq/geneseq/emb1/AA2001.DAT: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	63	68.5	534 18 AAW12844	RESULT 1 ID AAW12844 standard; Peptide: 534 AA.
2	63	68.5	535 18 AAW12841	XX AC AAW12844;
3	63	68.5	537 18 AAW12840	XX DT 15-DEC-1997 (first entry)
4	63	68.5	674 21 AAB53439	XX DE Pro-alpha2(I); (III)BGR chimeric protein.
5	63	68.5	1040 21 AAY84547	XX KW C-Propeptidase; recognition sequence; procollagen; monomer chain; therapy; KW trimersised pro-alpha chain; fibril; procollagen suicide; wound healing; KW fibrotic disease; human; chimeric protein.
6	63	68.5	1040 21 AAY84548	XX PA (UVM-) UNIV VICTORIA MANCHESTER.
7	63	68.5	1366 16 AAY1702	XX PT Bulleid N, Kadler K;
8	63	68.5	1366 21 AAY96123	XX DR WPT; 1997-179268/16.
9	63	68.5	1366 21 AAY56801	XX PT Novel pro:collagen mol. - comprising pro:collagen C-pro:peptide
10	63	68.5	1366 22 ABP50293	XX PT attached to an alien collagen alpha-chain or non-collagen material,
11	63	68.5	1366 22 AAE02536	XX PT Porcine alpha2(I)

ALIGNMENTS

OS	Home sapiens.	XX	XX	Type II collagen.
PN	W09708311-A1.	XX	PR	Human type II coll.
PD		XX	PR	Rat type II coll.
	06-MAR-1997.	XX	PR	Human cancer assoc
		XX	PR	Collagen alpha 1 (
		XX	PR	Collagen type III
		XX	PR	Human type II coll
		XX	PR	Collagen alpha 1 (
		XX	PR	Collagen type III
		XX	PR	Mouse type II coll
		XX	PR	Collagen alpha 1 (
		XX	PR	Collagen type III
		XX	PR	Mouse alpha 1 (XVI
		XX	PR	Collagen type III
		XX	PR	Bovine alpha1(III)
		XX	PR	Porcine alpha1(III)
		XX	PR	Novel human diagno
		XX	PR	Human collagen (IV
		XX	PR	Human colon cancer
		XX	PR	Amino acid sequenc
		XX	PR	Peptide #6891 enco
		XX	PR	Protein #6110 enco
		XX	PR	Human brain express
		XX	PR	Human bone marrow
		XX	PR	peptide #6937 enco
		XX	PR	Recombinant human
		XX	PR	Amino acid sequenc
		XX	PR	A C-terminal fragm

PT useful e.g. for wound healing
 XX
 PS Example 1; Page 38-41; 69pp; English.

This sequence represents a chimeric procollagen molecule of the invention. This sequence has the procollagen C-propeptide from the pro-alpha1(II) chain inserted into the pro-alpha2(I) chain sequence. The C-propeptide is implicated in the assembly of the monomer chains into trimerised pro-alpha chains prior to cleavage of the N- and C-propeptides and formation of collagen in fibril-forming pro-alpha chains. The C-propeptides determine the type-specific assembly of the moieties to which they are attached. The molecule of the invention comprises a first moiety having procollagen C-propeptide activity attached to a second moiety, which is an alien collagen alpha-chain or a non-collagen material. The novel collagen molecule can be used for treatment or diagnosis in humans or animals, especially for the treatment of procollagen suicide, as an adhesive or implant, to promote (chronic) wound healing or fibrotic diseases with reduced scarring or for use in photography, brewing, foodstuffs or textiles. The novel collagen molecules, especially when containing substitutions in the recognition site, may have significantly altered properties and characteristics, such as different binding kinetics or alpha-chain selection properties.

SQ Sequence 534 AA;
 Query Match 68.5%; Score 63; DB 18; Length 534;
 Best Local Similarity 73.3%; Pred. No. 0.65;
 Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 Qy 1 PAGPMPNGKDGKVG 15
 Db 223 PAGPSGPAGKDGRG 237

RESULT 2
 ID AAW12841
 ID AAW12841 standard; Protein; 535 AA.
 AC AAW12841;
 XX
 DT 15-DEC-1997 (first entry)
 XX Truncated pro-alpha2(I) chain.
 XX C-peptide; recognition sequence; procollagen; monomer chain; therapy;
 KW trimerised pro-alpha chain; fibril; procollagen suicide; wound healing;
 KW fibrotic disease; human.
 DE Homo sapiens.
 XX
 WO9708311-A1.
 XX
 PD 06-MAR-1997.
 XX
 PF 30-AUG-1996; 96WO-GB02122.
 XX
 PR 14-JUN-1996; 96GB-0012476.
 PR 31-AUG-1995; 95GB-0017773.
 PR 23-MAR-1996; 96GB-0006152.
 XX
 PA (UYMA-) UNIV VICTORIA MANCHESTER.
 XX
 PI Bulleid N, Kadler K;
 XX
 WPI; 1997-17928/16.
 DR N-PSDB; AAT59891.

XX Novel procollagen mol. - comprising pro:collagen C-pro:peptide attached to an alien collagen alpha-chain or non-collagen material, useful e.g. for wound healing

XX Example 1; Page 25-28; 69pp; English.

CC This sequence represents a truncated procollagen pro-alpha2(I) chain that can be used in the procollagen molecules of the invention. The C-propeptide is implicated in the assembly of the monomer chains into trimerised pro-alpha chains prior to cleavage of the N- and C-propeptides and formation of collagen in fibril-forming pro-alpha chains. The C-propeptides determine the type-specific assembly of the moieties to which they are attached. The molecule of the invention comprises a first moiety having procollagen C-propeptide activity attached to a second moiety, which is an alien collagen alpha-chain or a non-collagen material. The novel collagen molecule can be used for treatment or diagnosis in humans or animals, especially for the treatment of procollagen suicide, as an adhesive or implant, to promote (chronic) wound healing or fibrotic diseases with reduced scarring or for use in photography, brewing, foodstuffs or textiles. The novel collagen molecules, especially when containing substitutions in the recognition site, may have significantly altered properties and characteristics, such as different binding kinetics or alpha-chain selection properties.

SQ Sequence 535 AA;
 Query Match 68.5%; Score 63; DB 18; Length 535;
 Best Local Similarity 73.3%; Pred. No. 0.65;
 Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 Qy 1 PAGPMPNGKDGKVG 15
 Db 224 PAGPSGPAGKDGRG 238

RESULT 3
 ID AAW12840
 ID AAW12840 standard; peptide; 537 AA.
 AC AAW12840;
 XX
 DT 15-DEC-1997 (first entry)
 XX Pro-alpha2(I):(III) CP chimeric protein.
 XX C-propeptide; recognition sequence; procollagen; monomer chain; therapy;
 KW trimerised pro-alpha chain; fibril; procollagen suicide; wound healing;
 KW fibrotic disease; human; chimeric protein.
 OS Homo sapiens.
 XX
 PN WO9708311-A1:
 XX
 PD 06-MAR-1997.
 XX
 PF 30-AUG-1996; 96WO-GB02122.
 XX
 PR 14-JUN-1996; 96GB-0012476.
 PR 31-AUG-1995; 95GB-001773.
 PR 23-MAR-1996; 96GB-0006152.
 XX
 PA (UYMA-) UNIV VICTORIA MANCHESTER.
 XX
 PI Bulleid N, Kadler K;
 XX
 DR WPI; 1997-17928/16.

XX Novel pro:collagen mol. - comprising pro:collagen C-pro:peptide attached to an alien collagen alpha-chain or non-collagen material, useful e.g. for wound healing

XX Claim 16; Page 35-38; 69pp; English.

CC This sequence represents a chimeric procollagen molecule of the invention. This sequence has the procollagen C-prepropeptide from the pro-alpha1(II) chain attached to the pro-alpha2(I) chain sequence. The C-propeptide is implicated in the assembly of the monomer chains into trimerised proalpha chains prior to cleavage of the N- and C-propeptides and formation of collagen in fibril-forming pro-alpha chains. The

CC C-propeptides determine the type-specific assembly of the moieties to which they are attached. The molecule of the invention comprises a first CC moiety having procollagen C-propeptide activity attached to a second CC moiety, which is an alien collagen alpha-chain or a non-collagen material. The novel collagen molecule can be used for treatment or CC diagnosis in humans or animals, especially for the treatment of CC procollagen suicide, as an adhesive or implant, to promote (chronic) wound healing or fibrotic diseases with reduced scarring or for use in CC photography, brewing, foodstuffs or textiles. The novel collagen molecules, especially when containing substitutions in the recognition CC site, may have significantly altered properties and characteristics, such CC as different binding kinetics or alpha-chain selection properties.

XX SQ Sequence 537 AA;

Query Match 68.5%; Score 63; DB 18; Length 537;
Best Local Similarity 73.3%; Pred. No. 0.65;
Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Oy 1 PAGPMCPNGKDGKVG 15
Db 224 PAGPSGPGAKDGRTG 238

RESULT 4

ID AAB53439 standard; Protein: 674 AA.

XX AC AAB53439;

XX DT 09-MAR-2001 (first entry)

DE Human colon cancer antigen protein sequence SEQ ID NO:979.

KW Human; colon cancer; colon cancer antigen; diagnosis; detection; vulnerability; immunomodulatory; muscular; gynaecological; gastrointestinal; nephrotoxic; anti-infective; antibacterial; gene therapy; wound; neural disorder; immune system disorder; muscular disorder; reproductive disorder; gastrointestinal disorder; renal disorder; infectious disease; cardiovascular disorder.

XX OS Homo sapiens.

XX PN WO200055351-A1.

XX PD 21-SEP-2000.

XX PR 08-MAR-2000; 2000WO-US05983.

XX PR 12-MAR-1999; 99US1-0124270.

XX PA (HUMA-) HUMAN GENOME SCI INC.

PT Rosen CA, Ruben SM;

XX DR WPI; 2000-587534/55.

DR N-PSDB; AAC98195.

XX PT Colon cancer associated gene sequences, referred to as colon cancer

PT antigens, useful for the treatment, prevention, and diagnosis of colon disorders such as colon cancer -

Claim 11; Page 1551-1553; 210pp; English.

XX AAC97991 to AAC98763 encode the human colon cancer associated proteins, called human colon cancer antigens, given in AAB53434 to AAB54005. The human colon cancer antigens can have cytostatic, cardioactive, muscular, neuroprotective, immunomodulatory, gynaecological, gastrointestinal, vulnerable, nephrotoxic, anti-infective and antibacterial activities, and can be used in gene therapy. The colon cancer antigen polynucleotides, proteins and antibodies to the proteins are useful for the prevention, treatment and diagnosis of colon disorders, such as colon cancer. The

CC polyribonucleotides may be used in diagnostics and research, such as for CC chromosome identification, and as hybridisation probes. The proteins CC may also be used to prevent diseases such as neural disorders, immune CC system disorders, muscular disorders, reproductive disorders, gastrointestinal disorders, wounds, renal disorders, Infectious CC diseases, and cardiovascular disorders. AAC9764 to AAC9872 and CC AAC9807 represent sequences used in the exemplification of the present invention.

XX SQ Sequence 674 AA;

Query Match 68.5%; Score 63; DB 21; Length 674;
Best Local Similarity 73.3%; Pred. No. 0.82;
Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Oy 1 PAGPMCPNGKDGKVG 15
Db 363 PAGPSGPGAKDGRTG 377

RESULT 5

ID AAY84547 standard; Protein: 1040 AA.

XX AC AAY84547;

XX DT 25-JUL-2000 (first entry)

DE A human collagen 1 (alpha2) protein helical region.

KW Extracellular matrix protein; self aggregation; hydroxylated proline; trans-4-hydroxyproline; 3-hydroxyproline; recombinant protein production; collagen; fibrinogen; fibronectin; post translational hydroxylation.

XX OS Homo sapiens.

XX PN EP992586-A2.

XX PD 12-APR-2000.

XX PF 07-OCT-1999; 99EP-0119184.

XX PR 09-OCT-1998; 98US-0169768.

XX PA (USSU) US SURGICAL CORP.

XX PI Gruskin EA, Buechter DD, Zhang G, Connolly K;

XX DR WPI; 2000-259138/23.

XX DR N-PSDB; AAA12510.

XX PT Production of extracellular matrix proteins containing 4-trans-hydroxyproline results in native self aggregating proteins, useful on medical implants -

XX PS Example 14; Fig 49A-E; 260pp; English.

CC The specification describes a method for producing an extracellular CC matrix protein or its fragment. The extracellular matrix protein is CC capable of self aggregating in a cell which does not ordinarily CC hydroxylated prolines. The method comprises optimising a nucleic acid CC sequence for expression in the cell by substitution of codons preferred CC by that cell for naturally occurring codons not preferred by the cell; CC incorporating the nucleic acid sequence into the cell; and contacting CC the cell with a hypertonic growth medium containing at least one amino CC acid selected from the group consisting of trans-4-hydroxyproline and CC 3-hydroxyproline to allow at least one of the amino acids to be CC assimilated into the cell and incorporated into the extracellular matrix CC protein. The method may be used to make host cells assimilate and CC incorporate trans-4-hydroxyproline into proteins. This is especially CC useful in the recombinant production of proteins such as collagen, CC fibrinogen and fibronectin whose ability to self aggregate and produce CC functional proteins depends on the post-translational hydroxylation of

CC proline. The method is also useful in studying the structure and function of polypeptides which do not normally contain trans- α -hydroxyproline. The present sequence represents human collagen 1 (α_2) helical region, which may be produced using the method of the invention.

XX	Sequence	1040 AA:	CC	region' encoded by a sequence which has optimised codon usage. The protein may be produced using the method of the invention.
Query Match	68.5%; Score 63; DB 21; Length 1040;	CC		
Best Local Similarity	73.3%; Pred. No. 1.3;	CC		
Matches	11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;	CC		
OY	1 PAGPWGPNGPKDGKVG 15	CC		
Db	976 PAGPSGPAGKDGRIG 990	CC		
RESULT 6		Query Match	68.5%; Score 63; DB 21; Length 1040;	
ID AAY4548		Best Local Similarity	73.3%; Pred. No. 1.3;	
ID AAY4548 standard; Protein: 1040 AA.		Matches	11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;	
XX		OY	1 PAGPWGPNGPKDGKVG 15	
AC AAY4548;		Db	976 PAGPSGPAGKDGRIG 990	
XX		RESULT 7		
DT 25-JUL-2000 (first entry)		ID AAR1702		
XX		ID AAR1702 standard; protein: 1366 AA.		
DE A human collagen 1 (α_2) protein helical region.		XX		
XX		AC AAR1702;		
Extracellular matrix protein; self aggregation; hydroxylated proline; recombinant protein production; trans-4-hydroxyproline; 3-hydroxyproline; collagen; fibrinogen; fibronectin; post-translational hydroxylation.		XX		
KW		DT 17-OCT-1995 (first entry)		
KW		XX		
DE		DE Collagen alpha 2 (1) chain precursor.		
XX		XX		
OS Synthetic.		Collagen; antibody; immunoassay; metabolism; diagnosis; monitoring; disorder; osteoporosis; metastatic progression; Paget's disease; hyperthyroidism; bone; resorption; rheumatoid arthritis; osteoarthritis; vasculitis syndrome.		
OS Homo sapiens.		XX		
XX		OS Homo sapiens.		
PN EP992586-A2.		XX		
XX		PN WO9508115-A.		
PD 12-APR-2000.		XX		
XX		PD 23-MAR-1995.		
PP 07-OCT-1999; 99EP-0119184.		XX		
XX		PP 19-SEP-1994; 94WO-DK003148.		
PR 09-OCT-1998; 98US-0169768.		XX		
XX		PR 17-SEP-1993; 93DK-0001040.		
PA (OSTB-) OSTEOEMETER AS.		XX		
PA (USSU) US SURGICAL CORP.		PA Bonde M, Qvist P;		
XX		PI DR WPI; 1995-131456/17.		
PI Gruskin EA, Buechler DD, Zhang G, Connolly K;		XX		
XX		PT Assaying collagen fragments in body fluid by immunoassay - using antibodies raised against synthetic peptide(s) contg. potential crosslinking sites, to diagnose and monitor disorders of collagen metabolism, e.g. osteoporosis.		
DR WPI; 2000-259138/23.		PT Disclosure (Appendix A); Page 51; 8pp; English.		
DR N-PSDB; AAA12511.		PS Determination of collagen fragments in body fluids can be achieved by immunassay using antibodies directed against synthetic peptides derived from collagen which contain sites of potential crosslinking. The method is used to diagnose and monitor treatment of disorders of collagen metabolism (degradation of type I collagen may indicate osteoporosis, metastatic progression, Paget's disease, hyperthyroidism or other conditions involving excessive bone resorption; degradation of type II collagen may indicate rheumatoid arthritis or osteoarthritis; and of type III collagen, vasculitis (syndrome). The method can also be used to assess the toxicity of a compound and to test drugs for their effect on collagen metabolism.		
XX		XX Sequence 1366 AA;		
PT Production of extracellular matrix proteins containing 4-trans-hydroxyproline results in native self aggregating proteins, useful on medical implants -		Query Match 68.5%; Score 63; DB 16; Length 1366;		
PT Example 14; FIG 50A-E; 260pp; English.		Best Local Similarity 73.3%; Pred. No. 1.7;		
PS The specification describes a method for producing an extracellular matrix protein or its fragment. The extracellular matrix protein is capable of self aggregating in a cell which does not ordinarily hydroxylate prolines. The method comprises optimising a nucleic acid sequence for expression in the cell by substitution of codons preferred by that cell for naturally occurring codons not preferred by the cell; incorporating the nucleic acid sequence into the cell; and contacting the cell with a hypertonic growth medium containing at least one amino acid, selected from the group consisting of trans-4-hydroxyproline and 3-hydroxyproline to allow at least one of the amino acids to be assimilated into the cell and incorporated into the extracellular matrix protein. The method may be used to make host cells assimilate and incorporate trans-4-hydroxyproline into proteins. This is especially useful in the recombinant production of proteins such as collagen, fibrinogen and fibronectin whose ability to self aggregate and produce functional proteins depends on the post-translational hydroxylation of proline. The method is also useful in studying the structure and function of polypeptides which do not normally contain trans-4-hydroxyproline. The present sequence represents a human collagen 1 (α_2) helical		Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;		
CC		OY 1 PAGPWGPNGPKDGKVG 15		
CC		Db 1055 PAGPSGPAGKDGRIG 1069		

RESULT 8
 XX AAY96123 standard; Peptide: 1366 AA.
 ID AAY96123
 XX AAY96123;
 XX DT 19-DEC-2000 (first entry)
 XX DE Collagen type I alpha-2.
 XX KW Collagen type I; osteoporosis; bone resportion; Paget's disease; hyperparathyroidism; metastasis; assay; diagnosis.
 XX OS Homo sapiens.
 XX PN US6110689-A.
 XX PD 29-AUG-2000.
 XX PR 04-NOV-1997; 97US-0963825.
 XX PR 21-JAN-1994; 94US-0187319.
 XX PA (OSTE-) OSTROMETER AS.
 XX PT Bonde M, Qvist P;
 XX DR WPI; 2000-586349/55.
 XX Assaying type I collagen fragments for diagnosing osteoporosis in postmenopausal woman, involves contacting body fluid with synthetic collagen peptide and antibody and quantifying by competitive binding assay -
 XX PS disclosure; Column 31-37; 41pp; English.
 XX The present sequence is that of human type I collagen alpha-2.
 CC The invention is based on the discovery of the presence of particular collagen fragments in body fluids of patients compared with those of healthy subjects. These fragments are generated upon collagen degradation and are partly characterised by the presence of potential sites for crosslinking. A method for assaying collagen fragments in a body fluid sample is based on the competitive binding to immunological binding partners of collagen fragments in the sample and of synthetic peptides derived from collagen and containing crosslinkable sites (see AAY96105-11). When considering the degradation of type I collagen, the assay can be used as a means of identifying excessive bone resportion, indicating the presence of osteoporosis or the metastatic progres of a malignancy. Other conditions characterized by excessive bone resportion include Paget's disease and hyperparathyroidism.

SQ Sequence 1366 AA:

```
Query Match 68.5%; Score 63; DB 21; Length 1366;
Best Local Similarity 73.3%; Pred. No. 1.7; Length 1366;
Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
```

QY 1 PAGPWGNGKDKGVG 15
QY 1 PAGPWGNGKDKGVG 15
Db 1055 PAGPSGPAGKDERTG 1069

RESULT 9
 AAY56801 standard; protein: 1366 AA.
 ID AAY56801
 XX AC AAY56801;
 XX DT 27-MAR-2000 (first entry)
 XX XX Human preproalpha 2 (I) collagen.
 XX DE Fibrillar collagen; C propeptide; SSAD; telopeptide; gelatin; sequence selection and alignment domain; prosthetic implant; foodstuff; medicine; type I collagen; human.
 XX OS Homo sapiens.
 XX PN EB967226-A2.
 XX PD 29-DEC-1999.
 XX PF 04-MAY-1999; 99EP-0303470.
 XX PR 08-MAY-1998; 98US-0084828.
 PR 10-APR-1999; 99US-0289578.
 XX PA (COHE-) COHESION TECHNOLOGIES INC.
 XX Oiseen DR, Hitzeman RA, Chisholm GE;
 XX DR WPI; 2000-074666/07.
 XX PT New method for production of fibrillar collagen, useful for preparing telopeptide collagen fibrils and gelatin -
 XX PS Example 1; Fig 4A-B; 30pp; English.
 XX CC The invention provides a method for the production of fibrillar collagen. The method comprises: (a) culturing a recombinant host cell comprising a DNA encoding a fibrillar collagen monomer lacking a C propeptide SSAD (sequence selection and alignment domain); and (b) producing the fibrillar collagen. The methods are used to produce fibrillar collagen, from which telopeptide collagen fibrils can be derived. Host cells, comprising DNA encoding a collagen monomer lacking SSAD or N propeptide is used to produce gelatin. Collagen is used in biological research as a substrate for in vitro cell culture and as a component of biocompatible materials for use in prosthetic implants, sustained drug release matrices, artificial skin and wound dressing and healing devices. Gelatin is particularly useful for foodstuffs and medicine, for coating tablets and making capsules. The methods, comprising the use of collagen monomers lacking the N and/or C propeptides, result in a large increase in the production of type I collagen. The present sequence represents the human preproalpha 2 (I) collagen (GenBank Accn no: 274616).

SQ Sequence 1366 AA:

```
Query Match 68.5%; Score 63; DB 21; Length 1366;
Best Local Similarity 73.3%; Pred. No. 1.7; Length 1366;
Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
```

QY 1 PAGPWGNGKDKGVG 15
QY 1 PAGPWGNGKDKGVG 15
Db 1055 PAGPSGPAGKDERTG 1069

RESULT 10
 ABB5023 standard; Protein: 1366 AA.
 ID ABB50293
 XX AC ABB50293;
 XX ABB50293;
 XX DT 08-FEB-2002 (first entry)
 XX DE Collagen type I alpha-2 ovarian tumour marker protein, SEQ ID NO:76.
 XX KW Ovarian tumour marker gene; human; overexpression; upregulation; epithelial tumour; cancer; diagnosis; prognosis; disease monitoring; identification; serous cystadenoma; borderline serous tumour; serous cystadenocarcinoma; mucinous cystadenocarcinoma; mucinous adenocarcinoma; borderline mucinous tumour; endometrioid carcinoma; undifferentiated carcinoma; clear cell adenocarcinoma; cystadenofibroma; adenofibroma; Brenner tumour; serial analysis of gene expression; SAGE;

KW Collagen; triple helix; articular cartilage; collagenase;
 KW degradation; monoclonal antibody; epitope; matrix;
 KW metalloproteinase.
 OS Homo sapiens.
 XX
 PH Key Location/Qualifiers
 FT 1..24
 /label= Signal peptide.
 XX
 PN WO9414070-A.
 XX
 PD 23-JUN-1994.
 XX
 PT 04-DEC-1992; 92US-0984123.
 PR 17-JUL-1995; 95US-0448501.
 XX
 PA (SHRI-) SHRINERS HOSPITALS FOR CHILDREN.
 PI Billinghurst RC, Poole AR, Hollander AP;
 XX DR
 PN WPI; 2001-006136/01.
 XX
 PT Detecting cartilage degradation useful for early detection of arthritis
 PT or joint damage by contacting the biological sample with an antibody
 PT that binds to an epitope of unwound type II collagen chains but not to
 PT a native helical collagen -
 XX
 PS Example 1; Fig 1; 58pp; English.
 XX
 CC The present invention relates to detecting cartilage degradation in a
 CC biological sample by identifying the presence of unwound type II
 CC collagen in the sample. The method involves contacting the sample with
 CC a monoclonal antibody which only binds an epitope on unwound type II
 CC collagen chains. The invention is useful for the early detection of
 CC arthritis and joint damage and for monitoring disease related to
 CC collagen.
 XX SQ Sequence 1418 AA:
 Query Match 64.1%; Score 59; DB 22; Length 1418;
 Best Local Similarity 66.7%; Pred. No. 6.3.;
 Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
 QY 1 PAGPGPGNGKDGKVG 15
 OY | | | : | | | : | | | : |
 Db 1096 PPGPVGPGSKGDAMG 1110
 AC AARY9480;
 XX DT 19-JAN-1996 (first entry)
 XX DE Rat type II collagen.
 XX KW Collagen; bone progenitor; gene transfer; gene therapy; osteoporosis;
 KW osteotomy; bone repair; osteotropic; *Pichia pastoris*.
 XX OS Rattus sp.
 XX PN WO9522611-A2.
 XX PD 24-AUG-1995.
 XX
 AAB35624 ID AAB35624 standard; Protein; 1418 AA.
 AC AAB35624:
 XX DT 14-FEB-2001 (first entry)
 DE Human type II collagen.
 XX KW Type II collagen; arthritis; joint; ds.
 OS Homo sapiens.
 XX PN US6132976-A.
 PD 17-OCT-2000.
 XX
 PF 22-JAN-1998; 98US-0010999.
 XX
 PR 04-DEC-1992; 92US-0984123.
 PR 17-JUL-1995; 95US-0448501.
 XX
 PA (SHRI-) SHRINERS HOSPITALS FOR CHILDREN.
 PI Billinghurst RC, Poole AR, Hollander AP;
 XX DR
 PN WPI; 2001-006136/01.
 XX
 PT Detecting cartilage degradation useful for early detection of arthritis
 PT or joint damage by contacting the biological sample with an antibody
 PT that binds to an epitope of unwound type II collagen chains but not to
 PT a native helical collagen -
 XX
 PS Example 1; Fig 1; 58pp; English.
 XX
 CC The present invention relates to detecting cartilage degradation in a
 CC biological sample by identifying the presence of unwound type II
 CC collagen in the sample. The method involves contacting the sample with
 CC a monoclonal antibody which only binds an epitope on unwound type II
 CC collagen chains. The invention is useful for the early detection of
 CC arthritis and joint damage and for monitoring disease related to
 CC collagen.
 XX SQ Sequence 1418 AA:
 Query Match 64.1%; Score 59; DB 22; Length 1418;
 Best Local Similarity 66.7%; Pred. No. 6.3.;
 Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
 QY 1 PAGPGPGNGKDGKVG 15
 OY | | | : | | | : | | | : |
 Db 1096 PPGPVGPGSKGDAMG 1110
 AC AARY9480;
 XX DT 19-JAN-1996 (first entry)
 XX DE Rat type II collagen.
 XX KW Collagen; bone progenitor; gene transfer; gene therapy; osteoporosis;
 KW osteotomy; bone repair; osteotropic; *Pichia pastoris*.
 XX OS Rattus sp.
 XX PN WO9522611-A2.
 XX PD 24-AUG-1995.
 XX
 AAB35624 ID AAB35624 standard; Protein; 1418 AA.
 AC AAB35624:
 XX DT 14-FEB-2001 (first entry)
 DE Human type II collagen.
 XX KW Type II collagen; arthritis; joint; ds.
 OS Homo sapiens.
 XX PN US6132976-A.
 PD 17-OCT-2000.
 XX
 PF Transferring nucleic acid into bone progenitor cell(s) - using a
 PT bone compatible matrix, for treatment of fracture(s) and
 PT osteoporosis.
 XX
 PS Disclosure; Page 197-208; 317pp; English.
 XX
 CC Human, rat and mouse collagen type II (given in AAR79479-81).

Sequence 1442 AA: C
C respectively) can be used to stimulate bone progenitor cells as a means of treating bone-related diseases in association with an osteotropic gene.

rejection, modulate haemostatic or thrombolytic activity, modulate inflammation, cancers, cardiovascular disorders, neurological disease and bacterial or viral infections. The peptides, nucleotides, antibodies, agonists and antagonists may be also be used in drug screens. AAC18449 to AAC78457 and AAB4240 represent sequences used in the exemplification of the present invention.

Query Match 62.0%; Score 57; DB 16; Length 1442;
 Best Local Similarity 73.3%; Pred. NO. 12;
 Matches 11; Conservative 0; Mismatches 4;
 Matches 11; Conservative 0; Mismatches 4;
 Qy 1 PAGPWGPNGKDGKVG 15
 Db 871 PAGPWPAGKGDKPKG 885

XX
SQ Sequence 97 AA:

RESULT 15
NAB44187 ID:187 8+9999999. 07 27
Db 13 PVGPXGPPKGDTXG 27

RESULT 16
AR79481
AAR79481 standard; Protein; 54 AA

DE Human cancer associated protein sequence SEQ ID NO:1632.
XX Human; cancer associated gene; cancer antigen; detection; cancer;
KW diagnosis; cytostatic; proliferative; vulnerable; immunomodulator;
KW antidiabetic; antiasthmatic; antiheumatic; antiarthritic; antiviral;
KW antiinflammatory; antithyroid; antibacterial; cardiotonic;
KW dermatological; neuroprotective; thrombolytic; coagulant; nootropic;
KW vasoconstrictor; antipsoriatic; antiangiogenic; gene therapy; inflammation;
KW immune disorder; haematopoietic cell disorder; autoimmune disorder;
KW allergic reaction; graft versus host disease; organ rejection;
KW haemostatic; thrombolytic; cardiovascular disorder; infection;
KW neurological disease; drug screening;

24-AUG-1995 .
21-FEB-1995; 95WO-US02251

R
R
X
R
A

30-SEP-1994; 94US-0316650
18-FEB-1994; 94US-0199780

(UNMI) UNIV MICHIGAN.

R
T
transferring nucleic acid into bone progenitor cell(s) - using a
WPI; 1995-302717/39.

osteoporosis.
disclosure: Page 210; 317PP; English.

Human, rat and mouse collagen type II (given in AAR79479-81,

can be used to stimulate bone progenitor cells as a means of treating bone-related diseases in association with an osteotropic gene.

Sequence 54 AA;
Query Match 59.8%; Score 55; DB 16; Length 54;
Best Local Similarity 66.7%; Brad No 0.87;

CC AAC77607 to AAC78448 encode the human cancer associated proteins given in AAB4339 to AAB4439. The proteins can have activities based on the tissues and cells the genes are expressed in. Example of activities include: cytosatic; proliferative; pulmonary; immunomodulator; antidiabetic; antiasthmatic; antirheumatic; antiarthritic; antiinflammatory; antithyroid; antiallergic; antibacterial; antiviral; dermatological; neuroprotective; cardiot; thrombolytic; coagulant; nootropic; vasoconstrictor; antipsoriatic and antiangiogenic. The polynucleotides and polypeptides can be used for preventing, treating or ameliorating medical conditions and diagnosing pathological conditions. Polynucleotides, polypeptides, antibodies, agonists and antagonists from the present invention may be used to treat immune disorders by activating or inhibiting the proliferation, differentiation or mobilisation of immune cells, to treat disorders of haematopoietic cells, autoimmune disorders, allergic reactions, graft versus host disease and organ

XX
 DT 17-OCT-1995 (first entry)
 XX
 DE Collagen alpha 1 (II) chain precursor.
 XX
 KW Collagen; antibody; immunoassay; metabolism; diagnosis; monitoring;
 disorder; osteoporosis; metastatic progression; Paget's disease;
 KW hyperthyroidism; bone; resorption; rheumatoid arthritis;
 KW osteoarthritis; vasculitis syndrome.
 XX
 OS Homo sapiens.
 XX
 PN WO9508115-A.
 XX
 PD 23-MAR-1995.
 XX
 PF 19-SEP-1994; 94WO-DK00348.
 XX
 PR 17-SEP-1993; 93DK-0001040.
 XX
 PA (OSPE-) OSTEOmeter AS.
 XX
 PI Bonde M, Qvist P;
 XX
 DR WPI; 1995-131456/17.
 XX
 PT Assaying collagen fragments in body fluid by immunoassay - using
 antibodies raised against synthetic peptide(s) contg. Potential
 crosslinking sites, to diagnose and monitor disorders of collagen
 metabolism, e.g. osteoporosis.
 XX
 DR WPI; 1995-131456/17.
 XX
 PT Assaying collagen fragments in body fluid by immunoassay - using
 antibodies raised against synthetic peptide(s) contg. Potential
 crosslinking sites, to diagnose and monitor disorders of collagen
 metabolism, e.g. osteoporosis.
 XX
 PS Disclosure (Appendix A); Page 53; 87pp; English.
 XX
 CC Determination of collagen fragments in body fluids can be achieved
 by immunoassay using antibodies directed against synthetic peptides
 derived from collagen which contain sites of potential crosslinking.
 The method is used to diagnose and monitor treatment of disorders of
 collagen metabolism (degradation of type I collagen may indicate
 osteoporosis, metastatic progression, Paget's disease,
 hyperthyroidism or other conditions involving excessive bone
 resorption; degradation of type II collagen may indicate rheumatoid
 arthritis; or osteoarthritis; and of type III collagen, vacuilitis
 syndrome). The method can also be used to assess the toxicity of a
 compound and to test drugs for their effect on collagen metabolism.
 XX
 Sequence 1418 AA;
 SQ Query Match 59.8%; Score 55; DB 16; Length 1118;
 Best Local Similarity 66.7%; Pred. No. 23;
 Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
 ID AAV96124 standard; peptide: 1418 AA.
 OY 1 PADPWPNGPKDGKVG 15
 Db 1096 PPGIVGPGSKDGANG 1110
 RESULT 19
 RAW61562
 ID AAW61562 standard; protein: 1487 AA.
 AC AAW61562;
 XX
 DT 02-Nov-1998 (first entry)
 XX
 DE Human type II collagen alpha-chain protein.
 XX
 KW Type II collagen alpha-chain; human; immunoassay; antibody; C-terminal;
 epitope; cartilage; enzyme-linked immunosorbent assay; ELISA; therapy;
 KW diagnosis; arthritis; growth disorder; prognosis; drug screening;
 KW anti-arthritis agent; matrix metalloprotease inhibitor.
 XX
 OS Homo sapiens.
 XX
 PN WO935235-A1.
 XX
 PD 13-AUG-1998.
 XX
 PF 30-JAN-1998; 98WO-GB000304.
 XX
 PR 06-FEB-1997; 97GB-0002252.
 XX
 PA (UWSH-) UNIV SHEFFIELD.
 XX
 PI Croucher LJ, Hollander AP;
 XX
 PN US6110689-A.

DR WPI; 1998-447376/38.

XX Immunoassay kit containing two antibodies recognising coupled PT epitope(s) on collagen fragments - and new antibodies, for PT diagnosing arthritis etc., also prognosis and screening for PT anti-arthritis agents or inhibitors of matrix metallo-protease XX disclosure; Fig 2; 57pp; English.

This sequence represents the human type II collagen alpha-chain which is used in a method to produce an immunoassay kit comprising of two antibodies (Ab1 and Ab2), mono- or poly-clonal, or their fragments, that bind to two C-tiffree coupled epitopes (C-tiffree indicates any type II collagen fragment that is released from degraded cartilage). The kits are designed for sandwich immunoassays, specifically enzyme-linked immunosorbent assay (ELISA). The kits are used for therapy, diagnosis (e.g. serum or synovial fluid), and C-tiffree is systemic (present in urine, routine screening for arthritis and other cartilage diseases, also to diagnose growth disorders), prognosis (e.g. monitoring progression of rheumatoid arthritis and osteoarthritis, or monitoring treatment with growth hormone) and for drug screening (to identify, and assess efficacy of, anti-arthritis agents and matrix metalloprotease inhibitors). C-tiffree, derived from the N-terminus of the alpha 1 chain, have increased resistance to proteolysis, so can accumulate in vivo to a concentration that allows accurate measurement by immunoassay.

XX Sequence 1487 AA;

SQ Query Match 59.8%; Score 55; DB 19; Length 1487;

Best Local Similarity 66.7%; Pred. No. 24; Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 1 PAGPMWPNKGDKGVG 15
Db 1165 PPGPVPGPSKGDKANG 1179

PS Disclosure (Appendix A); Page 55; 87pp; English.

XX Determination of collagen fragments in body fluids can be achieved CC by immunoassay using antibodies directed against synthetic peptides CC derived from collagen which contain sites of potential crosslinking.

The method is used to diagnose and monitor treatment of disorders of CC collagen metabolism (degradation of type I collagen may indicate CC osteoporosis, metastatic progression, Paget's disease, CC hyperthyroidism or other conditions involving excessive bone CC resorption; degradation of type II collagen may indicate rheumatoid CC arthritis or osteoarthritis, and of type III collagen, vasculitis CC syndrome). The method can also be used to assess the toxicity of a CC compound and to test drugs for their effect on collagen metabolism.

XX Sequence 1078 AA;

SQ Query Match 58.7%; Score 54; DB 16; Length 1078;

Best Local Similarity 66.7%; Pred. No. 24; Matches 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 PAGPMWPNKGDKGVG 15
Db 994 PVGPSCGPPKGDTSG 1008

RESULT 21

AY96125
ID AAY96125 standard; Peptide; 1078 AA.

XX AC AAY96125;

XX DT 19-DEC-2000 (first entry)

XX DE Collagen type III alpha-1.

XX KW Collagen type III; vasculitis syndrome; assay; diagnosis.

XX OS Homo sapiens.

XX PN US6110689-A.

XX PD 29-AUG-2000.

XX PF 04-NOV-1997; 97US-0963825.

XX PR 21-JAN-1994; 94US-0187319.

XX PA (OSTE-) OSTEOMETER AS.

XX PI Bonde M, Qvist P;

XX DR WPI; 2000-586349/55.

PT Assaying type I collagen fragments for diagnosing osteoporosis in PT postmenopausal woman, involves contacting body fluid with synthetic PT collagen peptide and antibody and quantifying by competitive binding assay

XX Disclosure; Column 46-51; 41pp; English.

The present sequence is that of human type III collagen alpha-1. CC The invention is based on the discovery of the presence of the CC particular collagen fragments in body fluids of patients compared CC with those of healthy subjects. These fragments are generated CC upon collagen degradation and are partly characterised by the CC presence of potential sites for crosslinking. A method for assaying collagen fragments in a body fluid sample is based on the CC competitive binding to immunological binding partners of collagen CC fragments in the sample and of synthetic peptides derived from CC collagen and containing crosslinkable sites (see AAY96118-21). When CC considering the degradation of type III collagen, the assay can be CC used as a means of identifying the presence of vasculitis syndrome.

XX Assaying collagen fragments in body fluid by immunoassay - using PT antibodies raised against synthetic peptide(s) contg. potential PT crosslinking sites, to diagnose and monitor disorders of collagen PT metabolism, e.g. osteoporosis.

SQ	Sequence	1078 AA;	OS	Mus musculus.
Query Match	58.7%	Score 54; DB 21; Length 1078;	FH	location/Qualifiers
Best Local Similarity	66.7%	Pred. No. 24; 0; Mismatches	FT	303..308
Matches	10;	Conservative 5; Indels 0; Gaps 0;	FT	/label=GXYGX'Y'_motif
OY	1	PAGPWGPNGKDGKVG 15	FT	309..314
			Peptide	/label=GXYGX'Y'_motif
Db	994	PVPGSGPGKDGSQ 1008	FT	315..320
			Peptide	/label=GXYGX'Y'_motif
RESULT 22			FT	321..326
AAR28916			Peptide	/label=GXYGX'Y'_motif
ID	AAR28916	standard; Protein; 1196 AA.	FT	337..342
AC	AA28915;		Peptide	/label=GXYGX'Y'_motif
XX			FT	343..348
DT	24-MAR-1993	(first entry)	Peptide	/label=GXYGX'Y'_motif
DE	Type III procollagen (prior art).		FT	349..354
KW	Mutation; pro-alpha1(III); primer; PCR.		Peptide	/label=GXYGX'Y'_motif
XX	Homo sapiens.		FT	355..360
XX	W09219754-A.		Peptide	/label=GXYGX'Y'_motif
PN			FT	361..366
XX	12-NOV-1992.		Peptide	/label=GXYGX'Y'_motif
PD			FT	367..372
XX	08-MAY-1992;	92WO-US03866.	Peptide	/label=GXYGX'Y'_motif
XX	PR	08-MAY-1991; 91US-0696607.	FT	373..378
XX	(UVE-) UNIV JEFFERSON THOMAS.		Peptide	/label=GXYGX'Y'_motif
PA			FT	379..384
XX	Kuivaniemi SH, Prockop DJ, Tromp GC;		Peptide	/label=GXYGX'Y'_motif
PI			FT	385..390
XX	DR	WPI; 1992-399878/4B.	Peptide	/label=GXYGX'Y'_motif
NN-PSDB;	AAQ0849.		FT	396..401
XX			Peptide	/label=GXYGX'Y'_motif
PT	Kit for detecting genetic pre-disposition for vascular aneurysms		FT	402..407
PT	- contains primer to amplify portions of Type III procollagen DNA		Peptide	/label=GXYGX'Y'_motif
PS	Disclosure; Fig 1A-F; 44pp; English.		FT	435..440
XX	Example 1 describes the determination of the presence of a mutation		Peptide	/label=GXYGX'Y'_motif
CC	in the pro-alpha1(III) gene. Primers used in PCR are given in		FT	441..446
CC	AN03031-48.		Peptide	/label=GXYGX'Y'_motif
XX	Sequence 1196 AA;		FT	447..452
SQ			Peptide	/label=GXYGX'Y'_motif
Query Match	58.7%	Score 54; DB 13; Length 1196;	FT	453..458
Best Local Similarity	66.7%	Pred. No. 27; Matches 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0;	Peptide	/label=GXYGX'Y'_motif
OY	1	PAGPWGPNGKDGKVG 15	FT	459..464
			Peptide	/label=GXYGX'Y'_motif
Db	1141	PVPGSGPGKDGSQ 1155	FT	470..475
			Peptide	/label=GXYGX'Y'_motif
RESULT 23			FT	476..481
AAW26328			Peptide	/label=GXYGX'Y'_motif
ID	AAW26328	standard; Protein; 1288 AA.	FT	482..487
XX			Peptide	/label=GXYGX'Y'_motif
AC	AAW26328;		FT	488..493
XX			Peptide	/label=GXYGX'Y'_motif
DT	19-NOV-1997 (first entry)		FT	494..499
XX	Mouse alpha-1 collagen (XVII).		Peptide	/label=GXYGX'Y'_motif
XX	Alpha-1 collagen; type XVII collagen; cartilage degeneration.		FT	500..505
XX			Peptide	/label=GXYGX'Y'_motif
			FT	506..511
			Peptide	/label=GXYGX'Y'_motif
			FT	530..535
			Peptide	/label=GXYGX'Y'_motif
			FT	536..541
			Peptide	/label=GXYGX'Y'_motif
			FT	542..547
			Peptide	/label=GXYGX'Y'_motif
			FT	548..553
			Peptide	/label=GXYGX'Y'_motif
			FT	580..585
			Peptide	/label=GXYGX'Y'_motif

FT Peptide 586..591
 FT Peptide /label= GXYGX 'Y' _motif
 FT Peptide 592..597
 FT Peptide /label= GXYGX 'Y' _motif
 FT Peptide 598..603
 FT Peptide /label= GXYGX 'Y' _motif
 FT Peptide 604..609
 FT Peptide /label= GXYGX 'Y' _motif
 FT Peptide 610..615
 FT Peptide /label= GXYGX 'Y' _motif
 FT Peptide 616..621
 FT Peptide /label= GXYGX 'Y' _motif
 FT Peptide 622..627
 FT Peptide /label= GXYGX 'Y' _motif
 FT Peptide 628..633
 FT Peptide /label= GXYGX 'Y' _motif
 FT Peptide 634..639
 FT Peptide /label= GXYGX 'Y' _motif
 FT Peptide 640..665
 FT Peptide /label= GXYGX 'Y' _motif
 FT Peptide 657..662
 FT Peptide /label= GXYGX 'Y' _motif
 FT Peptide 677..682
 FT Peptide /label= GXYGX 'Y' _motif
 FT Peptide 683..688
 FT Peptide /label= GXYGX 'Y' _motif
 FT Peptide 689..694
 FT Peptide /label= GXYGX 'Y' _motif
 FT Peptide 695..700
 FT Peptide /label= GXYGX 'Y' _motif
 FT Peptide 707..712
 FT Peptide /label= GXYGX 'Y' _motif
 FT Peptide 713..718
 FT Peptide /label= GXYGX 'Y' _motif
 FT Peptide 735..740
 FT Peptide /label= GXYGX 'Y' _motif
 FT Peptide 741..746
 FT Peptide /label= GXYGX 'Y' _motif
 FT Peptide 747..752
 FT Peptide /label= GXYGX 'Y' _motif
 FT Peptide 759..764
 FT Peptide /label= GXYGX 'Y' _motif
 FT Peptide 765..770
 FT Peptide /label= GXYGX 'Y' _motif
 FT Peptide 771..776
 FT Peptide /label= GXYGX 'Y' _motif
 FT Peptide 787..792
 FT Peptide /label= GXYGX 'Y' _motif
 FT Peptide 793..798
 FT Peptide /label= GXYGX 'Y' _motif
 FT Peptide 799..804
 FT Peptide /label= GXYGX 'Y' _motif
 FT Peptide 815..820
 FT Peptide /label= GXYGX 'Y' _motif
 FT Peptide 821..826
 FT Peptide /label= GXYGX 'Y' _motif
 FT Peptide 827..832
 FT Peptide /label= GXYGX 'Y' _motif
 FT Peptide 833..838
 FT Peptide /label= GXYGX 'Y' _motif
 FT Peptide 839..844
 FT Peptide /label= GXYGX 'Y' _motif
 FT Peptide 845..850
 FT Peptide /label= GXYGX 'Y' _motif
 FT Peptide 863..868
 FT Peptide /label= GXYGX 'Y' _motif
 FT Peptide 869..874
 FT Peptide /label= GXYGX 'Y' _motif
 FT Peptide 875..880
 FT Peptide /label= GXYGX 'Y' _motif
 FT Peptide 891..896
 FT Peptide /label= GXYGX 'Y' _motif
 FT Peptide 897..902

FT Peptide 903..908
 FT Peptide /label= GXYGX 'Y' _motif
 FT Peptide 911..916
 FT Peptide /label= GXYGX 'Y' _motif
 FT Peptide 917..922
 FT Peptide /label= GXYGX 'Y' _motif
 FT Peptide 928..933
 FT Peptide /label= GXYGX 'Y' _motif
 FT Peptide 934..939
 FT Peptide /label= GXYGX 'Y' _motif
 FT Peptide 956..961
 FT Peptide /label= GXYGX 'Y' _motif
 FT Peptide 962..967
 FT Peptide /label= GXYGX 'Y' _motif
 FT Peptide 968..973
 FT Peptide /label= GXYGX 'Y' _motif
 FT Peptide 1126..1131
 FT Peptide /label= GXYGX 'Y' _motif
 FT Peptide 1145..1150
 FT Peptide /label= GXYGX 'Y' _motif
 FT Peptide 1193..1198
 FT Peptide /label= GXYGX 'Y' _motif

XX US5643783-A.
 XX PD 01-JUL-1997.
 XX PF 01-DEC-1993; 93US-0159784.
 XX PR 01-DEC-1993; 93US-0159784.
 XX PA (HARD) HARVARD COLLEGE.
 XX PI Oh SP, Olsen BR;
 XX DR N-PSDB; AAT84485.
 XX PT Nucleic acid encoding human alpha-1 collagen - for production of recombinant alpha-1 collagen, for use in the treatment of cartilage degeneration
 XX PS Disclosure; Fig 2; 35pp; English.
 Query Match 58.7%; Score 54; DB 18; Length 1288;
 Best local Similarity 66.7%; Fred. No. 29;
 Matches 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0
 QY 1 PAGPWGPNCRGKVG 15
 799 .. 804
 Db 341 PPGPQGPPGSKDGTG 355
 815..820
 821..826
 827..832
 833..838
 839..844
 845..850
 863..868
 869..874
 875..880
 891..896
 897..902

RESULT 24
 AAW92297
 ID AAW92297 standard; peptide; 1288 AA.
 XX
 AC AAW92297;
 XX
 DT 28-APR-1999 (first entry)
 XX
 DE Mouse alpha-1 (XVIII) collagen chain common sequence M018(common)28.
 XX Human; type XVIII collagen; liver disease; cirrhosis; detection;
 KW hepatocellular carcinoma; diagnosis.
 XX
 OS Mus sp.
 XX
 PN WO9856399-A1.
 XX
 PD 17-DEC-1998.
 XX

PF 12-JUN-1998; 9800-US12327.
 XX PR
 XX 12-JUN-1997; 97US-0049369.
 XX PA
 PA (FIFT-) ACAD FINLAND.
 PA (INRM) INST NAT SANTE & RECH MEDICALE.
 PA DR WPI; 1999-070292/06.
 PT XX
 PT Clement B, Pihlajaniemi T, Rehn M;
 PT XX
 PT DR
 XX PS Example 6; Fig 8; 56pp; English.
 XX A method has been developed for the detecting liver disease. The method comprises: (a) reacting a patient sample with antibodies (Ab) specific for collagen type XVIII (Col18); (b) measuring the amount of Ab-antigen complex (C) formed as indicator of the amount of Col18 present; (c) similar analysis of a non-diseased control; and (d) comparing the amounts of Col18 in the two samples to detect presence or progression of disease. Elevated levels of Col18 are: (i) indicative of disease, specifically cirrhosis; and (ii) predictive of the prognosis of disease, specifically hepatocellular carcinoma (there is a relationship between Col18 mRNA levels and tumour size and necrosis, and survival times are significantly higher in patients with higher Col18 levels). The method provides non-invasive, early and accurate diagnosis of liver disease. The present sequence represents the sequence common to mouse alpha-1 (XVIII) collagen chain from the present invention.

CC Sequence 1288 AA:

Query	Match	Score	Length	Pred.	No.	Mismatches	Indels	Gaps	CC
Qy	1 PAGPWGPNGKDGKVG 15	58.7%	54	DB	20	0	0	0	Best local similarity 66.7%; Pred. No. 29; Mismatches 10; Conservative 0; Indels 0; Gaps 0;
Db	341 PPGPQGPDKGPG 355								

RESULT 25

ABB50291 ID ABB50291 standard; Protein; 1466 AA.
 AC XX
 AC 08-FEB-2002 (first entry)
 DE Collagen type III alpha-1 ovarian tumour marker protein, SEQ ID NO:72.
 XX
 KW Ovarian tumour marker gene; human; overexpression; upregulation; epithelial tumour; cancer; diagnosis; disease monitoring; identification; serous cystadenoma; borderline serous tumour; mucinous cystadenoma; mucinous cystadenocarcinoma; undifferentiated carcinoma; clear cell adenocarcinoma; cystadenofibroma; adenofibroma; Brenner tumour; serial analysis of gene expression; SAGE; immune response pathway; cell proliferation regulation; protein folding; membrane localised; secreted; therapeutic target; cytostatic; gene therapy; vaccine.
 XX OS Homo sapiens.
 XX PN WO20017177-A2.
 XX PD 11-OCT-2001.
 XX PF 03-APR-2001; 2001WO-US10947.
 XX

PR 03-APR-2000; 2000US-194336P.
 XX PA
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PA Morin PJ, Sherman-Baust CA, Pizer ES, Hough CD;
 PA DR WPI; 2001-626450/72.
 PT XX
 PT Detecting and identifying ovarian tumor, identifying increased risk for developing ovarian cancer, and determining effectiveness of ovarian tumor marker gene -
 PT PS Claim 23; Page 114-117; 140pp; English.
 XX The invention relates to methods for diagnosing and prognosing ovarian tumours in an individual via the detection and measurement of the expression of ovarian tumour marker genes (ABA83081-ABA83122, ABAB3180, ABAB3182 and ABAB3184) or segments thereof (ABA83123-ABA83169, ABAB3179, ABAB83181 and ABAB3183). The methods of the invention are useful for detecting an ovarian tumour in a patient, for identifying an individual at increased risk for developing ovarian cancer, in prognostic tests for assessing the relative severity of ovarian cancer, in tests for monitoring a patient in remission from ovarian cancer and in tests for monitoring disease status in a patient being treated for ovarian cancer. The methods can additionally be used to identify a particular tumour as being an ovarian tumour (i.e., an epithelial ovarian tumour selected from serous cystadenoma, borderline serous tumour, serous cystadenocarcinoma, mucinous cystadenoma, borderline mucinous tumour, mucinous cystadenocarcinoma, endometrioid carcinoma, undifferentiated carcinoma, clear cell adenocarcinoma, cystadenofibroma, adenofibroma and Brenner tumour. The ovarian tumour marker genes of the invention were identified using SAGE (serial analysis of gene expression) and were found to be overexpressed in a broad variety of ovarian epithelial tumour cells relative to normal ovarian epithelial cells. The marker genes are implicated in immune response pathways, in the regulation of cell proliferation and in protein folding, and many of these are membrane-localised or secreted. In addition to their use as diagnostic and prognostic markers, the ovarian tumour marker genes or their encoded proteins may be used as therapeutic targets for the treatment and prevention of ovarian cancer. Sequences ABB50297-ABB50299 represent proteins encoded by ovarian tumour marker genes of the invention.

XX Sequence 1466 AA:

Query	Match	Score	Length	Pred.	No.	Mismatches	Indels	Gaps	CC
Qy	1 PAGPWGPNGKDGKVG 15	58.7%	54	DB	22	0	0	0	Best local similarity 66.7%; Pred. No. 33; Mismatches 10; Conservative 0; Indels 0; Gaps 0;
Db	1141 PVGPSSGPDKGTSG 1155								

RESULT 26

AAB02533 ID AAB02533 standard; Protein; 1466 AA.
 AC XX
 AC AAB02533:
 XX DT 10-AUG-2001 (first entry)
 DE Bovine alpha(I,II) collagen #1.
 XX KW Bovine; alpha(I,II) collagen; gelatin; cytostatic; viral infection; pharmaceutical; food industry; cosmetic; autoimmune disorder; vaccine; medical; arterial sealant; bone graft; dermal implant; haemostat; cancer; rheumatoid arthritis; beverage; photographic application.
 XX OS Bos sp.
 PN WO200134647-A2.

PD	17-MAY-2001.	XX
PF	10-NOV-2000; 2000WO-US30792.	XX
PR	12-NOV-1999; 99US-0439058.	XX
PR	10-NOV-2000; 2000US-0439058.	XX
PA	(FIBR-) FIBROGEN INC.	XX
PA	Bell MP, Neff TB, Polarek JW, Seeley TW;	XX
PT	WPI; 2001-335911/35.	XX
PT	Novel isolated and purified bovine or porcine collagens and gelatins useful in medical, pharmaceutical, food and cosmetic industries, as vaccine, and for treating autoimmune disorders, infections and cancer	-
PT	Novel isolated and purified bovine or porcine collagens and gelatins useful in medical, pharmaceutical, food and cosmetic industries, as vaccine, and for treating autoimmune disorders, infections and cancer	-
PS	Example 2; Fig 4; 168pp; English.	XX
PS	The present sequence is bovine alpha1(III) collagen. The present invention relates to recombinant synthesis of collagens and gelatins derived from animals. Collagen is useful in medical, pharmaceutical, food and cosmetic industries. Collagen is an important component of arterial sealants, bone grafts, drug delivery system, dermal implants, haemostats, and incontinence implants, and for treating autoimmune disorders such as rheumatoid arthritis. Collagen is useful in food products such as moisturisers. Recombinant gelatin is useful in vaccine formulations for treating viral infections, autoimmune diseases and cancer. Gelatin is useful in the manufacture or as a component of various pharmaceutical and medical devices and products, in food and beverage industries, in hair care and skin care products, as a glue or adhesive in various manufacturing processes, as a light-sensitive coating in various electronic devices, as photoresist base in photolithographic processes, in printing and photographic applications, in laboratory application, and as a component in various gels used for biochemical and electrophoretic analysis, including enzymographic gels.	XX
SQ	Sequence 1466 AA:	XX
Qy	Query Match 58.7%; Score 54; DB 22; Length 1466;	XX
Qy	Best Local Similarity 66.7%; Pred. No. 33; Matches 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0;	XX
Db	1141 PVGPVGPKDGKVG 15	XX
RESULT 27		XX
ID AAE02534	standard; Protein; 1466 AA.	XX
AC AAE02534;		XX
XX	10-AUG-2001 (first entry)	XX
DE Bovine alpha1(III) collagen #2.		XX
XX	Bovine; alpha1(III) collagen; gelatin; cytostatic; viral infection; pharmaceutical; food industry; cosmetic; autoimmune disorder; vaccine; medical; arterial sealant; bone graft; dermal implant; haemostat; cancer; rheumatoid arthritis; beverage; photographic application.	XX
OS Bos sp.		XX
PN WO200134647-A2.		XX
PD 17-MAY-2001.		XX
PF 10-NOV-2000; 2000WO-US30792.		XX
PR 12-NOV-1999; 99US-0439058.		XX
PR 10-NOV-2000; 2000US-0439058.		XX
PA (FIBR-) FIBROGEN INC.		XX
PA Bell MP, Neff TB, Polarek JW, Seeley TW;		XX
PT WPI; 2001-335911/35.		XX
PT Novel isolated and purified bovine or porcine collagens and gelatins useful in medical, pharmaceutical, food and cosmetic industries, as vaccine, and for treating autoimmune disorders, infections and cancer		-
PT Novel isolated and purified bovine or porcine collagens and gelatins useful in medical, pharmaceutical, food and cosmetic industries, as vaccine, and for treating autoimmune disorders, infections and cancer		-
PS Example 2; Fig 6; 168pp; English.		XX
PS The present sequence is bovine alpha1(III) collagen. The present invention relates to recombinant synthesis of collagens and gelatins derived from animals. Collagen is useful in medical, pharmaceutical, food and cosmetic industries. Collagen is an important component of arterial sealants, bone grafts, drug delivery system, dermal implants, haemostats, and incontinence implants, and for treating autoimmune disorders such as rheumatoid arthritis. Collagen is useful in food products such as sausages casings, and in cosmetics or facial and skin products such as moisturisers. Recombinant gelatin is useful in vaccine formulations for treating viral infections, autoimmune diseases and cancer. Gelatin is useful in the manufacture or as a component of various pharmaceutical and medical devices and products, in food and beverage industries, in hair care and skin care products, as a glue or adhesive in various manufacturing processes, as a light-sensitive coating in various electronic devices, as photoresist base in photolithographic processes, in printing and photographic applications, in laboratory application, and as a component in various gels used for biochemical and electrophoretic analysis, including enzymographic gels.	XX	
SQ Sequence 1466 AA:		XX
Qy Query Match 58.7%; Score 54; DB 22; Length 1466;		XX
Qy Best Local Similarity 66.7%; Pred. No. 33; Matches 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0;		XX
Db 1141 PVGPVGPKDGKVG 15		XX
RESULT 28		XX
ID AAE02537	standard; Protein; 1466 AA.	XX
AC AAE02537;		XX
XX	10-AUG-2001 (first entry)	XX
DE Porcine alpha1(III) collagen.		XX
XX	Porcine; alpha1(III) collagen; gelatin; cytostatic; viral infection; pharmaceutical; food industry; cosmetic; autoimmune disorder; vaccine; medical; arterial sealant; bone graft; dermal implant; haemostat; cancer; rheumatoid arthritis; beverage; photographic application.	XX
KW Sus scrofa.		XX
OS Sus scrofa.		XX
PN WO200134647-A2.		XX
PD 17-MAY-2001.		XX
PF 10-NOV-2000; 2000WO-US30792.		XX
PR 12-NOV-1999; 99US-0439058.		XX
PR 10-NOV-2000; 2000US-0439058.		XX

XX
PA (FIBR-) FIBROGEN INC.
XX
PT Bell MP, Neff TB, Polarek JW, Seeley TW;
XX
WPI: 2001-335911/35.
DR N-PSDB; AAD06578.

XX
PT Novel isolated and purified bovine or porcine collagens and gelatins useful in medical, pharmaceutical, food and cosmetic industries, as vaccine, and for treating autoimmune disorders, infections and cancer
XX
PS Example 5; Fig 12; 168pp; English.

XX
CC The present sequence is porcine alpha(III) collagen. The present invention relates to recombinant synthesis of collagens and gelatins derived from animals. Collagen is useful in medical, pharmaceutical, food and cosmetic industries. Collagen is an important component of arterial sealants, bone grafts, drug delivery system, dermal implants, haemostats, and incontinence implants, and for treating autoimmune disorders such as rheumatoid arthritis. Collagen is useful in food products such as sausage casings, and in cosmetics or facial and skin formulations for treating viral infections, autoimmune diseases and cancer. Gelatin is useful in the manufacture or as a component of various pharmaceutical and medical devices and products, in food and beverage industries, in hair care and skin care products, as a glue or adhesive in various manufacturing processes, as a light-sensitive coating in various electronic devices, as photoresist base in photolithographic processes, in printing and photographic applications, in laboratory application, and as a component in various gels used for biochemical and electrophoretic analysis, including enzymographic gels.
XX
SQ Sequence 1466 AA;

Query Match 58.7%; Score 54; DB 22; Length 146;
Best Local Similarity 66.7%; Pred. No. 33; Mismatches 0; Matches 10; Conservative 0; Indels 5; Gaps 0;

OY 1 PAGPWGPNGKCKVG 15
Db 1142 PVGPGSPPGKGASG 1156

RESULT 29

ABG15191
ID ABG15191 standard; Protein: 1469 AA.
XX
AC ABG15191;
XX
DT 18-FEB-2002 (first entry)

XX
DE Novel human diagnostic protein #15182.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic; food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PP 30-MAR-2001; 2001WO-US08631.
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
(HYSE-) HYSEQ INC.
XX
PT Drmanac RT, Liu C, Tang YT;

DR WPI: 2001-639362/73.
DR N-PSDB; AAST9378.

XX
PT New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity
XX
PS Claim 20; SEQ ID NO 45550; 103pp; English.

XX
CC The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, and gene mapping, and in recombinant production of (II). The poly nucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ABC01001-ABG0377 represent novel human diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp://wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1469 AA;

Query Match 58.7%; Score 54; DB 22; Length 1469;
Best Local Similarity 66.7%; Pred. No. 33; Mismatches 0; Matches 10; Conservative 0; Indels 5; Gaps 0;

OY 1 PAGPWGPNGKCKVG 15
Db 1144 PVGPGSPPGKGCTSG 1158

RESULT 30

AAR53257
ID AAR53257 standard; Protein: 1838 AA.
XX
AC AAR53257;
XX
DT 12-JAN-1995 (first entry)

XX
DE Human collagen (Type V).
XX
KW Human collagen; alpha 1; v type collagen; Placental mRNA.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Peptide 1..38
FT Region /label= signal_peptide
FT Binding-site 44..538
FT /note= "contains (Gly-X-Y) repeats"
FT /label= RGD
FT /note= "cell adhesion motif"
FT Binding-site 645..647
FT /label= RGD
FT /note= "heparin_binding_domain"
FT Domain 663..665
FT /label= C-terminal_region
FT /note= "contains 8 Cys residues"

XX
PN JP06105687-A.
XX
PD 19-APR-1994.
XX
PF 27-DEC-1991; 91JP-0358300.
PR 27-DEC-1991; 91JP-0358300.
XX
PA (TAKI) TAKARA SHUZO CO LTD.
XX
DR WPI; 1994-163129/20.
DR N-PSDB; AAQ64556.
XX
PT Human collagen V-type gene - is used for diagnosis of human
collagen V-type related diseases
XX
PS Claim 1; Page 6-14; 19PP; Japanese.
PT
XX
CC This amino acid sequence of type V collagen contains several distinct
domains including a region comprising repeated (Gly-X-Y) motifs and a
central domain containing two copies of the Arg-Gly-Asp cell adhesion
motif. The cDNA sequence encoding type V collagen was isolated from a
human placental library and will be useful for diagnosis of diseases
related to type V collagen.
CC
XX
SQ Sequence 1838 AA;

Query Match 58.7%; Score 54; DB 15; Length 1838;
Best Local Similarity 66.7%; Pred. No. 41;
Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
QY 1 PAGWGPNKGDKWIG 15
Db 1145 PAGWGPPGEDGDKC 1159

Search completed: November 1, 2002, 12:52:35
Job time : 28.5 secs

Copyright (c) 1993 - 2002 Compugen Ltd.

Gencore version 5.1.3

run on: November 1, 2002, 12:52:42 ; Search time 10.5 Seconds
 (without alignment)
 34.894 Million cell updates/sec

protein - protein search, using sw model

total number of hits satisfying chosen parameters: 231628

minimum DB seq length: 0

maximum DB seq length: 200000000

post-processing: Minimum Match 0%
 Maximum Match 100%
 Listing first 45 summaries

Issued_Patents_AA:*

1: /rgn2_6/ptodata/1/iaa/5A_COMB.pep:*

2: /rgn2_6/ptodata/1/iaa/5B_COMB.pep:*

3: /rgn2_6/ptodata/1/iaa/6A_COMB.pep:*

4: /rgn2_6/ptodata/1/iaa/6B_COMB.pep:*

5: /rgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*

6: /rgn2_6/ptodata/1/iaa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query	Length	DB	ID	Description
1	63	68.5	534	4	US-09-029-348-5	Sequence 5, Appli
2	63	68.5	535	4	US-09-029-348-1	Sequence 1, Appli
3	63	68.5	537	4	US-09-029-348-4	Sequence 4, Appli
4	63	68.5	1024	3	US-08-931-820-2	Sequence 2, Appli
5	63	68.5	1366	3	US-08-965-825-19	Sequence 1, Appli
6	63	68.5	1366	4	US-09-570-573-19	Sequence 19, Appli
7	63	68.5	1366	4	US-09-588-608-19	Sequence 1, Appli
8	59	64.1	1418	4	US-09-010-99-1	Sequence 1, Appli
9	57	62.0	1442	5	PCT-US95-02251-12	Sequence 12, Appli
10	55	59.8	41	1	US-08-935-816-2	Sequence 2, Appli
11	55	59.8	54	2	US-08-316-650-14	Sequence 14, Appli
12	55	59.8	54	5	PCT-US95-02251-14	Sequence 1, Appli
13	55	59.8	1060	3	US-08-931-820-3	Sequence 3, Appli
14	55	59.8	1418	3	US-08-933-823-20	Sequence 20, Appli
15	55	59.8	1418	4	US-09-570-573-20	Sequence 20, Appli
16	55	59.8	1418	4	US-09-570-573-20	Sequence 20, Appli
17	55	59.8	1418	4	US-09-548-608-20	Sequence 20, Appli
18	54	58.7	1057	3	US-08-931-820-4	Sequence 4, Appli
19	54	58.7	1078	3	US-08-965-825-21	Sequence 21, Appli
20	54	58.7	1078	4	US-09-570-573-21	Sequence 21, Appli
21	54	58.7	1078	4	US-09-548-608-21	Sequence 21, Appli
22	52	56.5	1057	3	US-08-931-820-1	Sequence 1, Appli
23	52	56.5	1341	3	US-08-963-825-18	Sequence 18, Appli
24	52	56.5	1341	4	US-09-570-573-18	Sequence 18, Appli
25	52	56.5	1341	4	US-09-588-608-18	Sequence 18, Appli
26	51	55.4	1057	1	US-07-951-565-1	Sequence 6, Appli
27	51	55.4	US-07-951-565-1			

ALIGNMENTS

Result 1	US-01-029-348-5	Sequence 5, Application US/09029348	Patent No. 617187	GENERAL INFORMATION:
APPLICANT: THE VICTORIA UNIVERSITY OF MANCHESTER				
TITLE OF INVENTION: NOVEL PROCOLLAGENS				
FILE REFERENCE: G08785PUS LISTING				
CURRENT APPLICATION NUMBER: US/09/029-348				
CURRENT FILING DATE: 1998-05-07				
NUMBER OF SEQ ID NOS: 20				
SOFTWARE: PatentIn Ver. 2.0				
SEQ ID NO 5 LENGTH: 534				
TYPE: PRT				
ORGANISM: Artificial Sequence				
FEATURE:				
OTHER INFORMATION: Description of Artificial Sequence: SEQUENCE				
OTHER INFORMATION: DERIVED FROM CDNA OF PROCOLLAGENS				
US-01-029-348-5				
Query Match 68.5%; Score 63; DB 4; Length 534;				
Best Local Similarity 73.3%; Ped. No. 018; Mismatches 3; Indels 0; Gaps 0				
QY	1	1 PAGPWGPNGKDGKV 15		
Db	223	PAGPSPGPKGDGRTG 237		
RESULT 2				
Sequence 1, Application US/09029348				
Patient No. 6171827				
GENERAL INFORMATION:				
APPLICANT: THE VICTORIA UNIVERSITY OF MANCHESTER				
TITLE OF INVENTION: NOVEL PROCOLLAGENS				
FILE REFERENCE: G08785PUS LISTING				
CURRENT APPLICATION NUMBER: US/09/029-348				
CURRENT FILING DATE: 1998-05-07				
NUMBER OF SEQ ID NOS: 20				
SOFTWARE: PatentIn Ver. 2.0				
SEQ ID NO 1 LENGTH: 535				
TYPE: PRT				
ORGANISM: Artificial Sequence				
FEATURE:				
OTHER INFORMATION: Description of Artificial Sequence: SEQUENCE				
OTHER INFORMATION: DERIVED FROM CDNA OF PROCOLLAGENS				
US-01-029-348-1				

Query Match 68.5%; Score 63; DB 4; Length 535;
 Best Local Similarity 73.3%; Pred. No. 0.18;
 Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 PAGPWNPGNGDGKVG 15
 Db 224 PAGPSGPAGKDGRG 238

RESULT 3 US-09-029-348-4

Sequence 4, Application US/09029348
 Patent No. 6171827

GENERAL INFORMATION:

APPLICANT: THE VICTORIA UNIVERSITY OF MANCHESTER
 TITLE OF INVENTION: A Novel Procollagens
 FILE REFERENCE: Q087857PUS LISTING

CURRENT APPLICATION NUMBER: US/09/029,348
 CURRENT FILING DATE: 1998-05-07
 NUMBER OF SEQ ID NOS: 20
 SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 4
 LENGTH: 537
 TYPE: PRT
 ORGANISM: Artificial sequence
 FEATURE:
 OTHER INFORMATION: Description of Artificial Sequence: SEQUENCE

US-09-029-348-4

Query Match 68.5%; Score 63; DB 4; Length 537;
 Best Local Similarity 73.3%; Pred. No. 0.18;
 Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 PAGPWNPGNGDGKVG 15
 Db 224 PAGPSGPAGKDGRG 238

RESULT 4 US-08-931-820-2

Sequence 2, Application US/08931820
 Patent No. 6010863

GENERAL INFORMATION:

APPLICANT:

TITLE OF INVENTION: Assay for collagen degradation

NUMBER OF SEQUENCES: 4

COMPUTER READABLE FORM:

COMPUTER TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-POS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/963,825
 FILING DATE: 21-JAN-1994
 CLASSIFICATION: 436
 PRIORITY APPLICATION DATA:
 APPLICATION NUMBER: US/08/187,319
 FILING DATE: 21-JAN-1994
 ATTORNEY / AGENT INFORMATION:
 NAME: Gogoris, Adda C
 REGISTRATION NUMBER: 29,714
 REFERENCE DOCKET NUMBER: 4305/08701
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 212-527-7700
 TELEFAX: 212-753-6237
 TELEX: 236637

INFORMATION FOR SEQ ID NO: 19:

SEQUENCE CHARACTERISTICS:
 LENGTH: 1365 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 ORIGINAL SOURCE:
 ORGANISM: Homo sapiens
 IMMEDIATE SOURCE:
 CLONE: collagen alpha 2- type I

US-08-963-825-19

Query Match 68.5%; Score 63; DB 3; Length 1366;
 Best Local Similarity 73.3%; Pred. No. 0.49;
 Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 PAGPWNPGNGDGKVG 15
 Db 1055 PAGPSGPAGKDGRG 1069

RESULT 5 US-08-963-825-19

Sequence 19, Application US/08963825
 Patent No. 6110689

GENERAL INFORMATION:

APPLICANT: Qvist, Per
 TITLE OF INVENTION: A Method for Assaying Collagen Fragments
 TITLE OF INVENTION: In Body Fluids, A Test Kit and Means for Carrying Out the
 TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
 TITLE OF INVENTION: Disorders Associated with the Metabolism of
 NUMBER OF SEQUENCES: 21

CORRESPONDENCE ADDRESS:

ADDRESSEE: Darby & Darby PC
 STREET: 805 Third Avenue
 CITY: New York
 STATE: New York
 COUNTRY: USA
 ZIP: 10022

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-POS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/963,825
 FILING DATE: 21-JAN-1994
 CLASSIFICATION: 436
 PRIORITY APPLICATION DATA:
 APPLICATION NUMBER: US/08/187,319
 FILING DATE: 21-JAN-1994
 ATTORNEY / AGENT INFORMATION:
 NAME: Gogoris, Adda C
 REGISTRATION NUMBER: 29,714
 REFERENCE DOCKET NUMBER: 4305/08701
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 212-527-7700
 TELEFAX: 212-753-6237
 TELEX: 236637

INFORMATION FOR SEQ ID NO: 19:

SEQUENCE CHARACTERISTICS:
 LENGTH: 1365 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 ORIGINAL SOURCE:
 ORGANISM: Homo sapiens
 IMMEDIATE SOURCE:
 CLONE: collagen alpha 2- type I

US-08-963-825-19

Query Match 68.5%; Score 63; DB 3; Length 1366;
 Best Local Similarity 73.3%; Pred. No. 0.49;
 Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 PAGPWNPGNGDGKVG 15
 Db 1055 PAGPSGPAGKDGRG 1069

RESULT 6 US-09-570-573-19

Sequence 19, Application US/09570573
 Patent No. 6342361

US-08-931-820-2

PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/984,123
 FILING DATE: 04-DEC-1992
 ATTORNEY/AGENT INFORMATION:
 NAME: Bent, Stephen A.
 REGISTRATION NUMBER: 29,768
 REFERENCE/DOCKET NUMBER: 032931/0212
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (202)672-5300
 TELEX: 904136-672-5399
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH 1418 amino acids
 TYPE: amino acid
 STRANDEDNESS:
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 ORIGINAL SOURCE:
 ORGANISM: Human Type II collagen
 US-09-010-999-1

Query Match 64.1%; Score 59; DB 4; Length 1418;
 Best Local Similarity 66.7%; Pred. No. 1.8; Matches 10;
 保守型 2; Mismatches 3; Indels 0; Gaps 0;
 Qy 1 PAGPWGPNGKDGKVG 15
 Db 1096 PPGPGVGPCKDGKVG 1110

RESULT 9
 US-08-316-650-12
 Sequence 12, Application US/08316650
 ;
 Patent No. 594296
 GENERAL INFORMATION:
 APPLICANT: Bonadio, Jeffrey
 APPLICANT: Roessler, Blake J.
 APPLICANT: Goldstein, Steven A.
 APPLICANT: Lin, Wushan
 TITLE OF INVENTION: METHODS AND COMPOSITIONS
 FOR STIMULATING BONE CELLS
 NUMBER OF SEQUENCES: 15
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Arnold, White & Durkee
 STREET: P.O. Box 4433
 CITY: Houston
 STATE: Texas
 COUNTRY: United States of America
 ZIP: 77210

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII
 SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US95/02251
 FILING DATE: CONCURRENTLY HEREWITH
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/316,650
 FILING DATE: 30-SEP-1994
 CLASSIFICATION:
 APPLICATION NUMBER: US 08/199,780
 FILING DATE: 18-FEB-1994
 CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:
 NAME: Parker, David L.
 REGISTRATION NUMBER: 32,165
 REFERENCE/DOCKET NUMBER: UMIC009P--

TELECOMMUNICATION INFORMATION:
 TELEPHONE: (512) 418-3000
 TELEX: (713) 789-2679
 FILING DATE: 30-SEP-1994
 CLASSIFICATION: 514

PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/199,780
 FILING DATE: 30-SEP-1994
 ATTORNEY/AGENT INFORMATION:
 NAME: Parker, David L.
 REGISTRATION NUMBER: 32,165
 REFERENCE/DOCKET NUMBER: UMIC:008
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (512) 418-3000
 TELEX: (713) 789-2679
 INFORMATION FOR SEQ ID NO: 12:
 SEQUENCE CHARACTERISTICS:
 LENGTH 1442 amino acids
 TYPE: amino acid
 STRANDEDNESS:
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 PCT-US95-02251-12

Query Match 62.0%; Score 57; DB 5; Length 1442;
 Best Local Similarity 73.3%; Pred. No. 3.6; Matches 11; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 Qy 1 PAGPWGPNGKDGKVG 15
 Db 871 PAGPPGPAGKDGPKG 885

RESULT 11
 US 08-395-816-2
 Sequence 2, Application US/08395816
 Patent No. 574751
 GENERAL INFORMATION:
 APPLICANT: Toshiki TANAKA et al.
 TITLE OF INVENTION: PEPTIDE DERIVATIVES HAVING
 TITLE OF INVENTION: BINDING ACTIVITY TO MODIFIED LOW DENSITY LIPOPROTEIN
 NUMBER OF SEQUENCES: 2
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Wenderoth, Lind & Ponack
 STREET: 805 Fifteenth Street, N.W., #700
 CITY: Washington
 STATE: D.C.
 COUNTRY: U.S.A.
 ZIP: 20005
 COMPUTER READABLE FORM:
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: MS-DOS
 SOFTWARE: Wordperfect 5.1
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/395,816
 APPLICATION DATE: FILING DATE: February 28, 1995
 CLASSIFICATION: 514
 ATTORNEY/AGENT INFORMATION:
 NAME: Warren M. Cheek, Jr.
 REGISTRATION NUMBER: 33,367
 REFERENCE/DOCKET NUMBER:
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 202-371-8850
 TELEFAX:
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 41 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 FEATURE:
 NAME/KEY: modified site
 LOCATION: 4, 7, 10, 13, 34, 37, 40
 IDENTIFICATION METHOD:
 OTHER INFORMATION: /note:- "4 Hyp"

Query Match 59.8%; Score 55; DB 1; length 41;
 Best Local Similarity 66.7%; Pred. No. 0.16; 5; Indels 0; Gaps 0;
 Matches 10; Conservative 0; Mismatches 5;

Qy 1 PAGGWPNGKDGKVG 15
 Db 6 PXGPGPKGKG 20

RESULT 12
 US 08-316-650-14
 Sequence 14, Application US/08316650
 GENERAL INFORMATION:
 APPLICANT:
 TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR STIMULATING BONE
 NUMBER OF SEQUENCES: 18
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Arnold, White & Durkee
 STREET: P.O. Box 4433
 CITY: Houston
 STATE: Texas
 COUNTRY: United States of America
 ZIP: 77210

COMPUTER READABLE FORM:
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US95/02251
 FILING DATE: CONCURRENTLY HEREWITH
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/316,650
 FILING DATE: 30-SEP-1994
 CLASSIFICATION:
 APPLICATION NUMBER: US 08/199,780
 FILING DATE: 18-FEB-1994
 CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:

COUNTRY: USA
 ZIP: 77210
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/316,650
 FILING DATE: 30-SEP-1994
 CLASSIFICATION: 514
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/199,780
 FILING DATE: 30-SEP-1994
 ATTORNEY/AGENT INFORMATION:
 NAME: Parker, David L.
 REGISTRATION NUMBER: 32,165
 REFERENCE/DOCKET NUMBER: UMIC:008
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (512) 418-3000
 TELEFAX: (713) 789-2679
 TELE: 79-0924
 INFORMATION FOR SEQ ID NO: 14:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 54 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: peptide

US-08-316-650-14

Query Match 59.8%; Score 55; DB 2; length 54;
 Best Local Similarity 66.7%; Pred. No. 0.22; 4; Indels 0; Gaps 0;
 Matches 10; Conservative 1; Mismatches 4;

Qy 1 PAGGWPNGKDGKVG 15
 Db 20 PPGPVGPSPKGANG 34

RESULT 13
 PCT-US95-02251-14
 Sequence 14, Application PC/TUS9502251
 GENERAL INFORMATION:
 APPLICANT:
 TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR STIMULATING BONE
 NUMBER OF SEQUENCES: 18
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Arnold, White & Durkee
 STREET: P.O. Box 4433
 CITY: Houston
 STATE: Texas
 COUNTRY: United States of America
 ZIP: 77210

COMPUTER READABLE FORM:
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US95/02251
 FILING DATE: CONCURRENTLY HEREWITH
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/316,650
 FILING DATE: 30-SEP-1994
 CLASSIFICATION:
 APPLICATION NUMBER: US 08/199,780
 FILING DATE: 18-FEB-1994
 CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:

NAME: Parker, David L.
 REGISTRATION NUMBER: 32,165
 REFERENCE/DOCKET NUMBER: UNIC009P--
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (512) 418-3000
 TELEX: 79-0224
 INFORMATION FOR SEQ ID NO: 14:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 54 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 PCT-US95-02251-14

Query Match 59.8%; Score 55; DB 5; Length 54;
 Best Local Similarity 66.7%; Pred. No. 0.22; 4; Indels 0; Gaps 0;
 Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 1 PAGPWNPGNGDKGKV 15
 Db 20 PPGPVGPGSGDGANG 34

RESULT 14
 US-08-931-820-3
 ; Sequence 3, Application US/08931820
 ; Patent No. 6010863
 ; GENERAL INFORMATION:
 ;
 ; APPLICANT:
 ; TITLE OF INVENTION: Assay for collagen degradation
 ; NUMBER OF SEQUENCES: 4
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatientIn Release #1.0, Version #1.25 (EPO)
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/931,820
 ; FILING DATE:
 ; CLASSIFICATION: 436
 ; PRIORITY APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/187,319
 ; FILING DATE: 21-JAN-1994
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Gogoris, Adda C
 ; REGISTRATION NUMBER: 29,714
 ; REFERENCE/DOCKET NUMBER: 4305/08701
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 212-527-7700
 ; TELEX: 236687
 ; INFORMATION FOR SEQ ID NO: 20:
 ;
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1418 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; ORIGINAL SOURCE:
 ; ORGANISM: Homo sapiens
 ; IMMEDIATE SOURCE:
 ; CLONE: COLLAGEN - ALPHA 1 (I)
 ; US-08-963-825-20

Query Match 59.8%; Score 55; DB 3; Length 1418;
 Best Local Similarity 66.7%; Pred. No. 6.7; 4; Indels 0; Gaps 0;
 Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 1 PAGPWNPGNGDKGKV 15
 Db 1096 PPGPVGPGSGDGANG 1110

RESULT 16
 US-09-570-573-20
 ; Sequence 20, Application US/09570573
 ; Patent No. 6342361
 ; GENERAL INFORMATION:
 ;
 ; APPLICANT: Qvist, Per
 ; APPLICANT: Bonde, Martin
 ; TITLE OF INVENTION: A Method for Assaying Collagen Fragments Carrying Out the
 ; TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for Carrying Out the
 ; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence or
 ; TITLE OF INVENTION: Disorders Associated with the Metabolism of
 ; NUMBER OF SEQUENCES: 21
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Darby & Darby PC
 ; STREET: 805 Third Avenue
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: USA
 ; ZIP: 10022

RESULT 15
 US-08-963-825-20
 ; Sequence 20, Application US/08933825
 ; Patent No. 6110689
 ; GENERAL INFORMATION:
 ;
 ; APPLICANT: Qvist, Per

CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/548,608
 FILING DATE: 21-JAN-1994
 CLASSIFICATION: 436
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/187,319
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: Gogoris, Adda C
 REGISTRATION NUMBER: 29,714
 REFERENCE/DOCKET NUMBER: 4305/08701
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 212-527-7700
 TELEX: 236687
 INFORMATION FOR SEQ ID NO: 21:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1078 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 ORIGINAL SOURCE: Homo sapiens
 IMMEDIATE SOURCE:
 CLONE: COLLAGEN ALPHA 1 (III)
 US-09-548-608-21

RESULT 23
 US-08-963-825-18
 Query Match 58.7%; Score 54; DB 4; Length 1078;
 Best Local Similarity 66.7%; Pred. No. 6.9; Mismatches 0; Indels 5; Gaps 0;
 Matches 10; Conservative 0; Misnmatches 5;

Qy	1	PAGPWGPNGKDGKVG	15
Db	925	:	939

Query Match 56.5%; Score 52; DB 3; Length 1341;
 Best Local Similarity 60.0%; Pred. No. 13; Mismatches 5; Indels 0; Gaps 0;
 Matches 9; Conservative 1; Misnmatches 5;

Qy	1	PAGPWGPNGKDGKVG	15
Db	962	:	976

RESULT 24
 US-09-570-573-18
 Query Match 56.5%; Score 52; DB 3; Length 1057;
 Best Local Similarity 60.0%; Pred. No. 13; Mismatches 5; Indels 0; Gaps 0;
 Matches 9; Conservative 1; Misnmatches 5;

TELEX: 236687
INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:
LENGTH: 26 amino acids

TYPE: AMINO ACID

TOPOLGY: linear

MOLECULE TYPE: peptide

HYPOTHETICAL: NO

FRAGMENT TYPE: internal

ORIGINAL SOURCE: Galius domesticus

ORGANISM: Galius domesticus

TISSUE TYPE: collagen

US-07-951-565-1

Query Match 55.4%; Score 51; DB 1; Length 26;
Best Local Similarity 53.3%; Pred. No. 0.37; Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 PAGPWGPNGKOCKVG 15
Db 1 PTGPLGPKGQIGELG 15

RESULT 27
US-07-951-565-6

Sequence 6, Application US/07951565

Patent No. 5309347

GENERAL INFORMATION:

APPLICANT: Trentham, David E.

ATTORNEY: Weiner, Howard L.

TITLE OF INVENTION: Method of Treating Rheumatoid Arthritis with

TITLE OF INVENTION: Type II Collagen

NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:

ADDRESSEE: Darby & Darby

STREET: 805 Third Ave.

CITY: New York

STATE: New York

COUNTRY: U.S.A.

ZIP: 10022-7513

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER SYSTEM: PC-DOS/MS-DOS

OPERATING SYSTEM: PC-DOS/MS-DOS

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/951,565

FILING DATE: 19920925

CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: Gogoris, Adda C.

REGISTRATION NUMBER: 29,714

REFERENCE/DOCKET NUMBER: 1010/07300

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 371-2600

TELEFAX: (202) 371-2540

INFORMATION FOR SEQ ID NO: 9:

SEQUENCE CHARACTERISTICS:
LENGTH: 26 amino acids

TYPE: amino acid

STRANDEDNESS: unknown

TOPOLGY: unknown

MOLECULE TYPE: protein

US-08-246-242-9

Query Match 55.4%; Score 51; DB 1; Length 26;

Best Local Similarity 53.3%; Pred. No. 0.37; Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 PAGPWGPNGKOCKVG 15
Db 1 PTGPLGPKGQIGELG 15

RESULT 28
US-08-246-242-9

; Sequence 9, Application US/08246242

; Patent No. 5675060

GENERAL INFORMATION:

APPLICANT: Benoist, C.

APPLICANT: Matliss, D.

APPLICANT: Koukoff, V.

TITLE OF INVENTION: Transgenic Arthritic Mice

NUMBER OF SEQUENCES: 9

CORRESPONDENCE ADDRESS:

ADDRESSEE: Sterne, Kessler, Goldstein & Fox

STREET: 1100 New York Avenue, Suite 600

CITY: Washington

STATE: D.C.

COUNTRY: U.S.A.

ZIP: 20005

COMPILER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/246,242

FILING DATE: 19-MAR-1994

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Goldstein, Jorge A.

REGISTRATION NUMBER: 29,021

REFERENCE/DOCKET NUMBER: 1383.0080000

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 371-2600

TELEFAX: (202) 371-2540

INFORMATION FOR SEQ ID NO: 9:

SEQUENCE CHARACTERISTICS:
LENGTH: 26 amino acids

TYPE: amino acid

STRANDEDNESS: unknown

TOPOLGY: unknown

MOLECULE TYPE: protein

US-08-246-242-9

Query Match 55.4%; Score 51; DB 1; Length 26;

Best Local Similarity 53.3%; Pred. No. 0.37; Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 PAGPWGPNGKOCKVG 15
Db 1 PTGPLGPKGQIGELG 15

RESULT 29
PCR-US86-00206-1

; Sequence 1, Application PC/TUS9600206

GENERAL INFORMATION:

APPLICANT: Immunologic Pharmaceutical Corporation

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR

TREATING RHEUMATOID ARTHRITIS

NUMBER OF SEQUENCES: 7

CORRESPONDENCE ADDRESS:

ADDRESSEE: Luppin & Kusmer

STREET: 200 State Street

CITY: Boston

STATE: MA

COUNTRY: USA

ZIP: 02109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US96/00206
 FILING DATE:
 CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:
 NAME: Kerner, Ann-Louise
 REGISTRATION NUMBER: 33,523
 REFERENCE/DOCKET NUMBER: IMZ-014PCT
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 617-466-6000
 TELEFAX: 617-466-6040
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 26 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 FRAGMENT TYPE: internal
 ORIGINAL SOURCE: Bos taurus type II collagen
 PCT-US96-00206-1

Query Match 55.4%; Score 51; DB 5; Length 26;
 Best Local Similarity 53.3%; Pred. No. 0.37;
 Matches 8; Conservative 3; Mismatches 4; Indels 0;
 Gaps 0;

Qy 1 PAGPWGPNGKDGKVG 15
 | ||| ||| :|:|
 Db 1 PTGPGLGPKGOTGEUG 15

LENGTH: 464 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: peptide

US-08-836-854-19
 Query Match 55.4%; Score 51; DB 2; Length 464;
 Best Local Similarity 66.7%; Pred. No. 7.6;
 Matches 10; Conservative 0; Mismatches 5; Indels 0;
 Gaps 0;
 Qy 1 PAGPWGPNGKDGKVG 15
 | ||| ||| :|:|
 Db 442 PKGPPGGPKDGLPG 456

Search completed: November 1, 2002, 12:54:45
 Job time: 11.5 secs

RESULT 30
 US-08-836-854-19
 Sequence 19 Application US/08836854
 Patent No. 5824547
 GENERAL INFORMATION:
 APPLICANT: HASHINO, Kinikazu
 APPLICANT: MATSUSHITA, Hideyuki
 APPLICANT: KATO, Ikunoshin
 TITLE OF INVENTION: METHOD OF PRODUCTION OF TRANSFECTED CELLS
 NUMBER OF SEQUENCES: 21
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Browdy and Neimark
 STREET: 419 Seventh Street N.W. Ste. 300
 CITY: Washington
 STATE: D.C.
 COUNTRY: USA
 ZTF: 20004
 COMPUTER READABLE FORM:
 MEDIUM TYPE: FLOPPY disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatientIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/836,854
 FILING DATE:
 CLASSIFICATION: 435
 PRTOR APPLICATION DATA:
 APPLICATION NUMBER: PCT/JP95/02425
 FILING DATE: 29-NOV-1995
 PRTOR APPLICATION DATA:
 APPLICATION NUMBER: 317721/1994
 FILING DATE: 29-NOV-1994
 ATTORNEY/AGENT INFORMATION:
 NAME: Browdy, Roger L.
 REGISTRATION NUMBER: 25,618
 REFERENCE/DOCKET NUMBER: HASHINO-1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (202) 628-5197
 TELEFAX: (202) 737-3528
 INFORMATION FOR SEQ ID NO: 19:
 SEQUENCE CHARACTERISTICS:

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model.

Run on: November 1, 2002, 12:38:32 ; Search time 13.5 Seconds

(without alignments) 106.766 Million cell updates/sec

Title: US-09-529-691a-3
Perfect score: 92
Sequence: 1 PAGPPGPNGKDGKVG 15

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched:

283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Maximum Match 0%

Listing first 45 summaries

Database : PIR 71:
1: pir1:*,
2: pir2:*,
3: pir3:*,
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No. Score Query Match Length DB ID Description

Result No.	Score	Query Match	Length	DB	ID	Description
1	69	75.0	460	2	T33110	hypothetical protein C18H7.3 - Caenorhabditis elegans
2	66	71.7	291	2	T20942	hypothetical prote
3	63	68.5	1366	1	CGHU25	hypothetical prote
4	60	65.2	1373	1	A43201	collagen alpha 2(I)
5	60	65.2	2944	2	A54849	collagen alpha 1(V)
6	57	62.0	365	2	S10847	collagen alpha 1(V)
7	57	62.0	1051	2	A35763	collagen alpha 2(I)
8	57	62.0	1419	2	A41182	collagen alpha 1(I)
9	57	62.0	1487	2	B41182	collagen alpha 1(I)
10	56	60.9	193	2	S07133	collagen alpha 1(I)
11	56	60.9	303	2	T1989	hypothetical prote
12	56	60.9	636	2	S41067	collagen alpha 1(I)
13	56	60.9	888	2	S28791	collagen alpha 1(X)
14	56	60.9	1075	2	T30842	serine-repeat anti
15	56	60.9	1486	1	B40333	collagen alpha 1(I)
16	56	60.9	1492	1	A40333	collagen alpha 1(I)
17	56	60.9	1805	1	CGHU26	collagen alpha 1(X)
18	55	59.8	172	2	D41132	collagen-related p
19	55	59.8	326	2	T16841	hypothetical prote
20	55	59.8	363	2	T16831	hypothetical prote
21	55	59.8	464	2	S59513	collagen II AI pro
22	55	59.8	615	2	A50259	collagen alpha 1(I)
23	55	59.8	964	1	CGCH25	collagen alpha 2(I)
24	55	59.8	1042	1	CGCH15	collagen alpha 1(I)
25	55	59.8	1418	2	T45467	collagen alpha 1(I)
26	55	59.8	1464	2	S59855	collagen alpha 1(I)
27	55	59.8	1487	1	CGHU6C	collagen alpha 1(I)
28	55	59.8	1546	2	CGHU2	collagen alpha 2(X)
29	58.7	228	1	CGHU2	collagen UC01 - p	

ALIGNMENTS

RESULT 1
T33110 hypothetical protein C18H7.3 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 21-Jan-2000
R;Tin-Wollam, A.; Fronick, W.
submitted to the EMBL Data Library, May 1998
A;Description: The sequence of C. elegans cosmid C18H7.
A;Reference number: 221284
A;Accession: T33110
A;Status: preliminary; translated from GB/EMBL/DDJB
A;Molecule type: DNA
A;Residues: 1-460 <TIN>
A;Cross-references: EMBL:AF067607; PIDN: AAC17641.1; GSPDB:GN00022; CESP:C18H7.3
A;Experimental source: strain Bristol N2; clone C18H7
C;Genetics:
A;Gene: CESP:C18H7.3
A;Map position: 4
A;Introns: B4/1
C;Superfamily: Phaseolus glycine-rich cell wall protein 1.8
Query Match 75.0%; Score 69; DB 2; Length 460;
Best Local Similarity 80.0%; Pred. No. 0.017; Mismatches 1; Indels 0; Gaps 0;
Matches 12; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 PAGPPGPNGKDGKVG 15
Db 237 PAGPPGPNGKDGKVG 251

RESULT 2
T20942 hypothetical protein F15A2.1 - caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C;Accession: T20942
R;Gregory, J.
submitted to the EMBL Data Library, March 1996
A;Reference number: 219349
A;Accession: T20942
A;Status: preliminary; translated from GB/EMBL/DDJB
A;Molecule type: DNA
A;Residues: 1-291 <WIL>
A;Cross-references: EMLI:Z70207; PIDN:CAA94128.1; GSPDB:GN00028; CBSP:F15A2.1
A;Experimental source: clone F15A2
C;Genetics:
A;Gene: CESP:F15A2.1
A;Map position: X
A;Introns: 50/3
C;Superfamily: unassigned collagens

Query Match 71.7%; Score 66; DB 2; Length 291;
Best Local Similarity 80.0%; Pred. No. 0.029; Mismatches 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 PAGPAGPAGKDGKVG 15
Db 183 PAGPAGPAGKDGKVG 197

RESULT 3

CGH2S collagen alpha 2(I) chain precursor - human
N;Alternate names: procollagen alpha 2(I) chain
C;Species: Homo sapiens (man)

C;Date: 30-Jun-1989 #sequence_revision 25-Aug-1995 #text_change 21-Jul-2000

C;Accession: A28500; S0824; S09176; I55311; A58111; A28472; R42165; A34405; A90567; I55306; A02865; J. Biol. Chem. 262, 16033-16036, 1987
A;Title: Organization of the human pro-alpha-2(I) collagen gene.

A;Reference number: A28500; MUID:88059862

A;Accession: A28500
A;Molecule type: mRNA
A;Residues: I¹-V¹⁸, N¹⁹-D²¹-P²²-I²³-G²⁴-A²⁵-R²⁶-T²⁷-P²⁸-A²⁹-G³⁰-P³¹-K³²-G³³-D³⁴-A³⁵-V³⁶-I³⁷, V³⁸-A³⁹-G⁴⁰-R⁴¹-A⁴²-P⁴³-A⁴⁴-G⁴⁵-V⁴⁶-A⁴⁷-R⁴⁸-P⁴⁹-A⁵⁰-T⁵¹-G⁵²-A⁵³-V⁵⁴-I⁵⁵-A⁵⁶-R⁵⁷-P⁵⁸-A⁵⁹-G⁶⁰-V⁶¹-A⁶²-R⁶³-P⁶⁴-A⁶⁵-G⁶⁶-V⁶⁷-A⁶⁸-R⁶⁹-P⁷⁰-A⁷¹-G⁷²-V⁷³-A⁷⁴-R⁷⁵-P⁷⁶-A⁷⁷-G⁷⁸-V⁷⁹-A⁸⁰-R⁸¹-P⁸²-A⁸³-G⁸⁴-V⁸⁵-A⁸⁶-R⁸⁷-P⁸⁸-A⁸⁹-G⁹⁰-V⁹¹-A⁹²-R⁹³-P⁹⁴-A⁹⁵-G⁹⁶-V⁹⁷-A⁹⁸-R⁹⁹-P¹⁰⁰-A¹⁰¹-G¹⁰²-V¹⁰³-A¹⁰⁴-R¹⁰⁵-P¹⁰⁶-A¹⁰⁷-G¹⁰⁸-V¹⁰⁹-A¹¹⁰-R¹¹¹-P¹¹²-A¹¹³-G¹¹⁴-V¹¹⁵-A¹¹⁶-R¹¹⁷-P¹¹⁸-A¹¹⁹-G¹²⁰-V¹²¹-A¹²²-R¹²³-P¹²⁴-A¹²⁵-G¹²⁶-V¹²⁷-A¹²⁸-R¹²⁹-P¹³⁰-A¹³¹-G¹³²-V¹³³-A¹³⁴-R¹³⁵-P¹³⁶-A¹³⁷-G¹³⁸-V¹³⁹-A¹⁴⁰-R¹⁴¹-P¹⁴²-A¹⁴³-G¹⁴⁴-V¹⁴⁵-A¹⁴⁶-R¹⁴⁷-P¹⁴⁸-A¹⁴⁹-G¹⁵⁰-V¹⁵¹-A¹⁵²-R¹⁵³-P¹⁵⁴-A¹⁵⁵-G¹⁵⁶-V¹⁵⁷-A¹⁵⁸-R¹⁵⁹-P¹⁶⁰-A¹⁶¹-G¹⁶²-V¹⁶³-A¹⁶⁴-R¹⁶⁵-P¹⁶⁶-A¹⁶⁷-G¹⁶⁸-V¹⁶⁹-A¹⁷⁰-R¹⁷¹-P¹⁷²-A¹⁷³-G¹⁷⁴-V¹⁷⁵-A¹⁷⁶-R¹⁷⁷-P¹⁷⁸-A¹⁷⁹-G¹⁸⁰-V¹⁸¹-A¹⁸²-R¹⁸³-P¹⁸⁴-A¹⁸⁵-G¹⁸⁶-V¹⁸⁷-A¹⁸⁸-R¹⁸⁹-P¹⁹⁰-A¹⁹¹-G¹⁹²-V¹⁹³-A¹⁹⁴-R¹⁹⁵-P¹⁹⁶-A¹⁹⁷-G¹⁹⁸-V¹⁹⁹-A²⁰⁰-R²⁰¹-P²⁰²-A²⁰³-G²⁰⁴-V²⁰⁵-A²⁰⁶-R²⁰⁷-P²⁰⁸-A²⁰⁹-G²¹⁰-V²¹¹-A²¹²-R²¹³-P²¹⁴-A²¹⁵-G²¹⁶-V²¹⁷-A²¹⁸-R²¹⁹-P²²⁰-A²²¹-G²²²-V²²³-A²²⁴-R²²⁵-P²²⁶-A²²⁷-G²²⁸-V²²⁹-A²³⁰-R²³¹-P²³²-A²³³-G²³⁴-V²³⁵-A²³⁶-R²³⁷-P²³⁸-A²³⁹-G²⁴⁰-V²⁴¹-A²⁴²-R²⁴³-P²⁴⁴-A²⁴⁵-G²⁴⁶-V²⁴⁷-A²⁴⁸-R²⁴⁹-P²⁵⁰-A²⁵¹-G²⁵²-V²⁵³-A²⁵⁴-R²⁵⁵-P²⁵⁶-A²⁵⁷-G²⁵⁸-V²⁵⁹-A²⁶⁰-R²⁶¹-P²⁶²-A²⁶³-G²⁶⁴-V²⁶⁵-A²⁶⁶-R²⁶⁷-P²⁶⁸-A²⁶⁹-G²⁷⁰-V²⁷¹-A²⁷²-R²⁷³-P²⁷⁴-A²⁷⁵-G²⁷⁶-V²⁷⁷-A²⁷⁸-R²⁷⁹-P²⁸⁰-A²⁸¹-G²⁸²-V²⁸³-A²⁸⁴-R²⁸⁵-P²⁸⁶-A²⁸⁷-G²⁸⁸-V²⁸⁹-A²⁹⁰-R²⁹¹-P²⁹²-A²⁹³-G²⁹⁴-V²⁹⁵-A²⁹⁶-R²⁹⁷-P²⁹⁸-A²⁹⁹-G³⁰⁰-V³⁰¹-A³⁰²-R³⁰³-P³⁰⁴-A³⁰⁵-G³⁰⁶-V³⁰⁷-A³⁰⁸-R³⁰⁹-P³¹⁰-A³¹¹-G³¹²-V³¹³-A³¹⁴-R³¹⁵-P³¹⁶-A³¹⁷-G³¹⁸-V³¹⁹-A³²⁰-R³²¹-P³²²-A³²³-G³²⁴-V³²⁵-A³²⁶-R³²⁷-P³²⁸-A³²⁹-G³³⁰-V³³¹-A³³²-R³³³-P³³⁴-A³³⁵-G³³⁶-V³³⁷-A³³⁸-R³³⁹-P³⁴⁰-A³⁴¹-G³⁴²-V³⁴³-A³⁴⁴-R³⁴⁵-P³⁴⁶-A³⁴⁷-G³⁴⁸-V³⁴⁹-A³⁵⁰-R³⁵¹-P³⁵²-A³⁵³-G³⁵⁴-V³⁵⁵-A³⁵⁶-R³⁵⁷-P³⁵⁸-A³⁵⁹-G³⁶⁰-V³⁶¹-A³⁶²-R³⁶³-P³⁶⁴-A³⁶⁵-G³⁶⁶-V³⁶⁷-A³⁶⁸-R³⁶⁹-P³⁷⁰-A³⁷¹-G³⁷²-V³⁷³-A³⁷⁴-R³⁷⁵-P³⁷⁶-A³⁷⁷-G³⁷⁸-V³⁷⁹-A³⁸⁰-R³⁸¹-P³⁸²-A³⁸³-G³⁸⁴-V³⁸⁵-A³⁸⁶-R³⁸⁷-P³⁸⁸-A³⁸⁹-G³⁹⁰-V³⁹¹-A³⁹²-R³⁹³-P³⁹⁴-A³⁹⁵-G³⁹⁶-V³⁹⁷-A³⁹⁸-R³⁹⁹-P⁴⁰⁰-A⁴⁰¹-G⁴⁰²-V⁴⁰³-A⁴⁰⁴-R⁴⁰⁵-P⁴⁰⁶-A⁴⁰⁷-G⁴⁰⁸-V⁴⁰⁹-A⁴¹⁰-R⁴¹¹-P⁴¹²-A⁴¹³-G⁴¹⁴-V⁴¹⁵-A⁴¹⁶-R⁴¹⁷-P⁴¹⁸-A⁴¹⁹-G⁴²⁰-V⁴²¹-A⁴²²-R⁴²³-P⁴²⁴-A⁴²⁵-G⁴²⁶-V⁴²⁷-A⁴²⁸-R⁴²⁹-P⁴³⁰-A⁴³¹-G⁴³²-V⁴³³-A⁴³⁴-R⁴³⁵-P⁴³⁶-A⁴³⁷-G⁴³⁸-V⁴³⁹-A⁴⁴⁰-R⁴⁴¹-P⁴⁴²-A⁴⁴³-G⁴⁴⁴-V⁴⁴⁵-A⁴⁴⁶-R⁴⁴⁷-P⁴⁴⁸-A⁴⁴⁹-G⁴⁵⁰-V⁴⁵¹-A⁴⁵²-R⁴⁵³-P⁴⁵⁴-A⁴⁵⁵-G⁴⁵⁶-V⁴⁵⁷-A⁴⁵⁸-R⁴⁵⁹-P⁴⁶⁰-A⁴⁶¹-G⁴⁶²-V⁴⁶³-A⁴⁶⁴-R⁴⁶⁵-P⁴⁶⁶-A⁴⁶⁷-G⁴⁶⁸-V⁴⁶⁹-A⁴⁷⁰-R⁴⁷¹-P⁴⁷²-A⁴⁷³-G⁴⁷⁴-V⁴⁷⁵-A⁴⁷⁶-R⁴⁷⁷-P⁴⁷⁸-A⁴⁷⁹-G⁴⁸⁰-V⁴⁸¹-A⁴⁸²-R⁴⁸³-P⁴⁸⁴-A⁴⁸⁵-G⁴⁸⁶-V⁴⁸⁷-A⁴⁸⁸-R⁴⁸⁹-P⁴⁹⁰-A⁴⁹¹-G⁴⁹²-V⁴⁹³-A⁴⁹⁴-R⁴⁹⁵-P⁴⁹⁶-A⁴⁹⁷-G⁴⁹⁸-V⁴⁹⁹-A⁵⁰⁰-R⁵⁰¹-P⁵⁰²-A⁵⁰³-G⁵⁰⁴-V⁵⁰⁵-A⁵⁰⁶-R⁵⁰⁷-P⁵⁰⁸-A⁵⁰⁹-G⁵¹⁰-V⁵¹¹-A⁵¹²-R⁵¹³-P⁵¹⁴-A⁵¹⁵-G⁵¹⁶-V⁵¹⁷-A⁵¹⁸-R⁵¹⁹-P⁵²⁰-A⁵²¹-G⁵²²-V⁵²³-A⁵²⁴-R⁵²⁵-P⁵²⁶-A⁵²⁷-G⁵²⁸-V⁵²⁹-A⁵³⁰-R⁵³¹-P⁵³²-A⁵³³-G⁵³⁴-V⁵³⁵-A⁵³⁶-R⁵³⁷-P⁵³⁸-A⁵³⁹-G⁵⁴⁰-V⁵⁴¹-A⁵⁴²-R⁵⁴³-P⁵⁴⁴-A⁵⁴⁵-G⁵⁴⁶-V⁵⁴⁷-A⁵⁴⁸-R⁵⁴⁹-P⁵⁵⁰-A⁵⁵¹-G⁵⁵²-V⁵⁵³-A⁵⁵⁴-R⁵⁵⁵-P⁵⁵⁶-A⁵⁵⁷-G⁵⁵⁸-V⁵⁵⁹-A⁵⁶⁰-R⁵⁶¹-P⁵⁶²-A⁵⁶³-G⁵⁶⁴-V⁵⁶⁵-A⁵⁶⁶-R⁵⁶⁷-P⁵⁶⁸-A⁵⁶⁹-G⁵⁷⁰-V⁵⁷¹-A⁵⁷²-R⁵⁷³-P⁵⁷⁴-A⁵⁷⁵-G⁵⁷⁶-V⁵⁷⁷-A⁵⁷⁸-R⁵⁷⁹-P⁵⁸⁰-A⁵⁸¹-G⁵⁸²-V⁵⁸³-A⁵⁸⁴-R⁵⁸⁵-P⁵⁸⁶-A⁵⁸⁷-G⁵⁸⁸-V⁵⁸⁹-A⁵⁹⁰-R⁵⁹¹-P⁵⁹²-A⁵⁹³-G⁵⁹⁴-V⁵⁹⁵-A⁵⁹⁶-R⁵⁹⁷-P⁵⁹⁸-A⁵⁹⁹-G⁶⁰⁰-V⁶⁰¹-A⁶⁰²-R⁶⁰³-P⁶⁰⁴-A⁶⁰⁵-G⁶⁰⁶-V⁶⁰⁷-A⁶⁰⁸-R⁶⁰⁹-P⁶¹⁰-A⁶¹¹-G⁶¹²-V⁶¹³-A⁶¹⁴-R⁶¹⁵-P⁶¹⁶-A⁶¹⁷-G⁶¹⁸-V⁶¹⁹-A⁶²⁰-R⁶²¹-P⁶²²-A⁶²³-G⁶²⁴-V⁶²⁵-A⁶²⁶-R⁶²⁷-P⁶²⁸-A⁶²⁹-G⁶³⁰-V⁶³¹-A⁶³²-R⁶³³-P⁶³⁴-A⁶³⁵-G⁶³⁶-V⁶³⁷-A⁶³⁸-R⁶³⁹-P⁶⁴⁰-A⁶⁴¹-G⁶⁴²-V⁶⁴³-A⁶⁴⁴-R⁶⁴⁵-P⁶⁴⁶-A⁶⁴⁷-G⁶⁴⁸-V⁶⁴⁹-A⁶⁵⁰-R⁶⁵¹-P⁶⁵²-A⁶⁵³-G⁶⁵⁴-V⁶⁵⁵-A⁶⁵⁶-R⁶⁵⁷-P⁶⁵⁸-A⁶⁵⁹-G⁶⁶⁰-V⁶⁶¹-A⁶⁶²-R⁶⁶³-P⁶⁶⁴-A⁶⁶⁵-G⁶⁶⁶-V⁶⁶⁷-A⁶⁶⁸-R⁶⁶⁹-P⁶⁷⁰-A⁶⁷¹-G⁶⁷²-V⁶⁷³-A⁶⁷⁴-R⁶⁷⁵-P⁶⁷⁶-A⁶⁷⁷-G⁶⁷⁸-V⁶⁷⁹-A⁶⁸⁰-R⁶⁸¹-P⁶⁸²-A⁶⁸³-G⁶⁸⁴-V⁶⁸⁵-A⁶⁸⁶-R⁶⁸⁷-P⁶⁸⁸-A⁶⁸⁹-G⁶⁹⁰-V⁶⁹¹-A⁶⁹²-R⁶⁹³-P⁶⁹⁴-A⁶⁹⁵-G⁶⁹⁶-V⁶⁹⁷-A⁶⁹⁸-R⁶⁹⁹-P⁷⁰⁰-A⁷⁰¹-G⁷⁰²-V⁷⁰³-A⁷⁰⁴-R⁷⁰⁵-P⁷⁰⁶-A⁷⁰⁷-G⁷⁰⁸-V⁷⁰⁹-A⁷¹⁰-R⁷¹¹-P⁷¹²-A⁷¹³-G⁷¹⁴-V⁷¹⁵-A⁷¹⁶-R⁷¹⁷-P⁷¹⁸-A⁷¹⁹-G⁷²⁰-V⁷²¹-A⁷²²-R⁷²³-P⁷²⁴-A⁷²⁵-G⁷²⁶-V⁷²⁷-A⁷²⁸-R⁷²⁹-P⁷³⁰-A⁷³¹-G⁷³²-V⁷³³-A⁷³⁴-R⁷³⁵-P⁷³⁶-A⁷³⁷-G⁷³⁸-V⁷³⁹-A⁷⁴⁰-R⁷⁴¹-P⁷⁴²-A⁷⁴³-G⁷⁴⁴-V⁷⁴⁵-A⁷⁴⁶-R⁷⁴⁷-P⁷⁴⁸-A⁷⁴⁹-G⁷⁵⁰-V⁷⁵¹-A⁷⁵²-R⁷⁵³-P⁷⁵⁴-A⁷⁵⁵-G⁷⁵⁶-V⁷⁵⁷-A⁷⁵⁸-R⁷⁵⁹-P⁷⁶⁰-A⁷⁶¹-G⁷⁶²-V⁷⁶³-A⁷⁶⁴-R⁷⁶⁵-P⁷⁶⁶-A⁷⁶⁷-G⁷⁶⁸-V⁷⁶⁹-A⁷⁷⁰-R⁷⁷¹-P⁷⁷²-A⁷⁷³-G⁷⁷⁴-V⁷⁷⁵-A⁷⁷⁶-R⁷⁷⁷-P⁷⁷⁸-A⁷⁷⁹-G⁷⁸⁰-V⁷⁸¹-A⁷⁸²-R⁷⁸³-P⁷⁸⁴-A⁷⁸⁵-G⁷⁸⁶-V⁷⁸⁷-A⁷⁸⁸-R⁷⁸⁹-P⁷⁹⁰-A⁷⁹¹-G⁷⁹²-V⁷⁹³-A⁷⁹⁴-R⁷⁹⁵-P⁷⁹⁶-A⁷⁹⁷-G⁷⁹⁸-V⁷⁹⁹-A⁸⁰⁰-R⁸⁰¹-P⁸⁰²-A⁸⁰³-G⁸⁰⁴-V⁸⁰⁵-A⁸⁰⁶-R⁸⁰⁷-P⁸⁰⁸-A⁸⁰⁹-G⁸¹⁰-V⁸¹¹-A⁸¹²-R⁸¹³-P⁸¹⁴-A⁸¹⁵-G⁸¹⁶-V⁸¹⁷-A⁸¹⁸-R⁸¹⁹-P⁸²⁰-A⁸²¹-G⁸²²-V⁸²³-A⁸²⁴-R⁸²⁵-P⁸²⁶-A⁸²⁷-G⁸²⁸-V⁸²⁹-A⁸³⁰-R⁸³¹-P⁸³²-A⁸³³-G⁸³⁴-V⁸³⁵-A⁸³⁶-R⁸³⁷-P⁸³⁸-A⁸³⁹-G⁸⁴⁰-V⁸⁴¹-A⁸⁴²-R⁸⁴³-P⁸⁴⁴-A⁸⁴⁵-G⁸⁴⁶-V⁸⁴⁷-A⁸⁴⁸-R⁸⁴⁹-P⁸⁵⁰-A⁸⁵¹-G⁸⁵²-V⁸⁵³-A⁸⁵⁴-R⁸⁵⁵-P⁸⁵⁶-A⁸⁵⁷-G⁸⁵⁸-V⁸⁵⁹-A⁸⁶⁰-R⁸⁶¹-P⁸⁶²-A⁸⁶³-G⁸⁶⁴-V⁸⁶⁵-A⁸⁶⁶-R⁸⁶⁷-P⁸⁶⁸-A⁸⁶⁹-G⁸⁷⁰-V⁸⁷¹-A⁸⁷²-R⁸⁷³-P⁸⁷⁴-A⁸⁷⁵-G⁸⁷⁶-V⁸⁷⁷-A⁸⁷⁸-R⁸⁷⁹-P⁸⁸⁰-A⁸⁸¹-G⁸⁸²-V⁸⁸³-A⁸⁸⁴-R⁸⁸⁵-P⁸⁸⁶-A⁸⁸⁷-G⁸⁸⁸-V⁸⁸⁹-A⁸⁹⁰-R⁸⁹¹-P⁸⁹²-A⁸⁹³-G⁸⁹⁴-V⁸⁹⁵-A⁸⁹⁶-R⁸⁹⁷-P⁸⁹⁸-A⁸⁹⁹-G⁹⁰⁰-V⁹⁰¹-A⁹⁰²-R⁹⁰³-P⁹⁰⁴-A⁹⁰⁵-G⁹⁰⁶-V⁹⁰⁷-A⁹⁰⁸-R⁹⁰⁹-P⁹¹⁰-A⁹¹¹-G⁹¹²-V⁹¹³-A⁹¹⁴-R⁹¹⁵-P⁹¹⁶-A⁹¹⁷-G⁹¹⁸-V⁹¹⁹-A⁹²⁰-R⁹²¹-P⁹²²-A⁹²³-G⁹²⁴-V⁹²⁵-A⁹²⁶-R⁹²⁷-P⁹²⁸-A⁹²⁹-G⁹³⁰-V⁹³¹-A⁹³²-R⁹³³-P⁹³⁴-A⁹³⁵-G⁹³⁶-V⁹³⁷-A⁹³⁸-R⁹³⁹-P⁹⁴⁰-A⁹⁴¹-G⁹⁴²-V⁹⁴³-A⁹⁴⁴-R⁹⁴⁵-P⁹⁴⁶-A⁹⁴⁷-G⁹⁴⁸-V⁹⁴⁹-A⁹⁵⁰-R⁹⁵¹-P⁹⁵²-A⁹⁵³-G⁹⁵⁴-V⁹⁵⁵-A⁹⁵⁶-R⁹⁵⁷-P⁹⁵⁸-A⁹⁵⁹-G⁹⁶⁰-V⁹⁶¹-A⁹⁶²-R⁹⁶³-P⁹⁶⁴-A⁹⁶⁵-G⁹⁶⁶-V⁹⁶⁷-A⁹⁶⁸-R⁹⁶⁹-P⁹⁷⁰-A⁹⁷¹-G⁹⁷²-V⁹⁷³-A⁹⁷⁴-R⁹⁷⁵-P⁹⁷⁶-A⁹⁷⁷-G⁹⁷⁸-V⁹⁷⁹-A⁹⁸⁰-R⁹⁸¹-P⁹⁸²-A⁹⁸³-G⁹⁸⁴-V⁹⁸⁵-A⁹⁸⁶-R⁹⁸⁷-P⁹⁸⁸-A⁹⁸⁹-G⁹⁹⁰-V⁹⁹¹-A⁹⁹²-R⁹⁹³-P⁹⁹⁴-A⁹⁹⁵-G⁹⁹⁶-V⁹⁹⁷-A⁹⁹⁸-R⁹⁹⁹-P¹⁰⁰⁰-A¹⁰⁰¹-G¹⁰⁰²-V¹⁰⁰³-A¹⁰⁰⁴-R¹⁰⁰⁵-P¹⁰⁰⁶-A¹⁰⁰⁷-G¹⁰⁰⁸-V¹⁰⁰⁹-A¹⁰¹⁰-R¹⁰¹¹-P¹⁰¹²-A¹⁰¹³-G¹⁰¹⁴-V¹⁰¹⁵-A¹⁰¹⁶-R¹⁰¹⁷-P¹⁰¹⁸-A¹⁰¹⁹-G¹⁰²⁰-V¹⁰²¹-A¹⁰²²-R¹⁰²³-P¹⁰²⁴-A¹⁰²⁵-G¹⁰²⁶-V¹⁰²⁷-A¹⁰²⁸-R¹⁰²⁹-P¹⁰³⁰-A¹⁰³¹-G¹⁰³²-V¹⁰³³-A¹⁰³⁴-R¹⁰³⁵-P¹⁰³⁶-A¹⁰³⁷-G¹⁰³⁸-V¹⁰³⁹-A¹⁰⁴⁰-R¹⁰⁴¹-P¹⁰⁴²-A¹⁰⁴³-G¹⁰⁴⁴-V¹⁰⁴⁵-A¹⁰⁴⁶-R¹⁰⁴⁷-P¹⁰⁴⁸-A¹⁰⁴⁹-G¹⁰⁵⁰-V¹⁰⁵¹-A¹⁰⁵²-R¹⁰⁵³-P¹⁰⁵⁴-A¹⁰⁵⁵-G¹⁰⁵⁶-V¹⁰⁵⁷-A¹⁰⁵⁸-R¹⁰⁵⁹-P¹⁰⁶⁰-A¹⁰⁶¹-G¹⁰⁶²-V¹⁰⁶³-A¹⁰⁶⁴-R¹⁰⁶⁵-P¹⁰⁶⁶-A¹⁰⁶⁷-G¹⁰⁶⁸-V¹⁰⁶⁹-A¹⁰⁷⁰-R¹⁰⁷¹-P¹⁰⁷²-A¹⁰⁷³-G¹⁰⁷⁴-V¹⁰⁷⁵-A¹⁰⁷⁶-R¹⁰⁷⁷-P¹⁰⁷⁸-A¹⁰⁷⁹-G¹⁰⁸⁰-V¹⁰⁸¹-A¹⁰⁸²-R¹⁰⁸³-P¹⁰⁸⁴-A¹⁰⁸⁵-G¹⁰⁸⁶-V¹⁰⁸⁷-A¹⁰⁸⁸-R¹⁰⁸⁹-P¹⁰⁹⁰-A¹⁰⁹¹-G¹⁰⁹²-V¹⁰⁹³-A¹⁰⁹⁴-R¹⁰⁹⁵-P¹⁰⁹⁶-A¹⁰⁹⁷-G¹⁰⁹⁸-V¹⁰⁹⁹-A<sup

Best Local Similarity 73.3%; Pred. No. 0.38; Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;	
QY 1 PAGPWGPNGDKRGV 15	Db 1055 PAGPSGPGAKDGRTG 1069
RESULT 4	A43391
A;Cross-references: GB:168663; NUID:95187/161	A;Cross-references: GB:168663; NUID:95187/161
A;Status: preliminary; translated from GB/EMBL/DDBJ	A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA	A;Molecule type: mRNA
A;Residues: 663-675, 'V', 677, 'P', 679-742, 'A', 744-746 <FOR>	A;Residues: 663-675, 'V', 677, 'P', 679-742, 'A', 744-746 <FOR>
A;Title: Incorporation of type I collagen molecules that contain a mutant alpha 2(I) chain into extracellular matrix	A;Title: Incorporation of type I collagen molecules that contain a mutant alpha 2(I) chain into extracellular matrix
A;Reference number: 155369; NUID:93054637	A;Reference number: 155369; NUID:93054637
A;Status: translated from GB/EMBL/DDBJ	A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA	A;Molecule type: mRNA
A;Residues: 665-666, 'D', 668-670 <NY>	A;Residues: 665-666, 'D', 668-670 <NY>
A;Cross-references: GB:1.00613; NID:9180888; PIDN:AAB59384_1; PID:9180889	A;Cross-references: GB:1.00613; NID:9180888; PIDN:AAB59384_1; PID:9180889
A;Note: mutant sequence from a patient with osteogenesis imperfecta	A;Note: mutant sequence from a patient with osteogenesis imperfecta
A;Accession: A56799	A;Accession: A56799
R;Battman, J.F.; Hannanen, M.; Chan, D.; Cole, W.G.	R;Battman, J.F.; Hannanen, M.; Chan, D.; Cole, W.G.
Biochem. J. 276, 765-770, 1991	Biochem. J. 276, 765-770, 1991
A;Title: Characterization of a type I collagen alpha 2(I) glycine-586 to valine substitution	A;Title: Characterization of a type I collagen alpha 2(I) glycine-586 to valine substitution
A;Method: mutagenesis	A;Method: mutagenesis
A;Reference number: A56799; NUID:91291136	A;Reference number: A56799; NUID:91291136
A;Accession: A56799	A;Accession: A56799
A;Molecule type: mRNA	A;Molecule type: mRNA
A;Residues: 672-675, 'V', 677, 'P', 679-681 <BAR>	A;Residues: 672-675, 'V', 677, 'P', 679-681 <BAR>
A;Cross-references: GB:539878; NID:9179911; PIDN:AAPI9314_1; PID:9232761	A;Cross-references: GB:539878; NID:9179911; PIDN:AAPI9314_1; PID:9232761
A;Note: sequence extracted from NCBI Backbone (NCBInet) database	A;Note: sequence extracted from NCBI Backbone (NCBInet) database
A;Note: mutant sequence of patient with osteogenesis imperfecta type IV; the authors suggest control sequence	A;Note: mutant sequence of patient with osteogenesis imperfecta type IV; the authors suggest control sequence
R;Maekelae, J.K.; Vuorio, T.; Vuorio, E.	R;Maekelae, J.K.; Vuorio, T.; Vuorio, E.
Biochim. Biophys. Acta 1049, 171-176, 1990	Biochim. Biophys. Acta 1049, 171-176, 1990
A;Title: Growth-dependent modulation of type I collagen production and mRNA levels in cultured fibroblasts	A;Title: Growth-dependent modulation of type I collagen production and mRNA levels in cultured fibroblasts
A;Reference number: S10768; NUID:90304220	A;Reference number: S10768; NUID:90304220
A;Molecule type: mRNA	A;Molecule type: mRNA
A;Accession: S10768	A;Accession: S10768
A;Residues: 960-1021, 'L', 1023-1188, 'D', 1190-1197, 'S', 1199-1356 <MAE>	A;Residues: 960-1021, 'L', 1023-1188, 'D', 1190-1197, 'S', 1199-1356 <MAE>
A;Cross-references: EMBL:X55255; NID:930101; PIDN:CAA39142_1; PID:930102	A;Cross-references: EMBL:X55255; NID:930101; PIDN:CAA39142_1; PID:930102
A;Experimental source: fibroblast cell culture	A;Experimental source: fibroblast cell culture
R;Myers, J.C.; Chu, M.L.; Faro, S.H.; Clark, W.J.; Prockop, D.J.; Ramirez, F.	R;Myers, J.C.; Chu, M.L.; Faro, S.H.; Clark, W.J.; Prockop, D.J.; Ramirez, F.
Proc. Natl. Acad. Sci. U.S.A. 78, 3516-3520, 1981	Proc. Natl. Acad. Sci. U.S.A. 78, 3516-3520, 1981
A;Title: Cloning a cDNA for the pro-alpha2(I) chain of human type I collagen.	A;Title: Cloning a cDNA for the pro-alpha2(I) chain of human type I collagen.
A;Reference number: A18855; NUID:81273090	A;Reference number: A18855; NUID:81273090
A;Molecule type: mRNA	A;Molecule type: mRNA
A;Residues: 964-979, 'V', 981-1018, 'Q', 1020 <NYE>	A;Residues: 964-979, 'V', 981-1018, 'Q', 1020 <NYE>
A;Cross-references: GB:J00114; NID:9180393; PIDN:AAA51996_1; PID:9180394	A;Cross-references: GB:J00114; NID:9180393; PIDN:AAA51996_1; PID:9180394
A;Note: 1019-Leu was also found	A;Note: 1019-Leu was also found
R;Wenstrup, R.J.; Cohn, D.H.; Cohen, T.; Byers, P.H.	R;Wenstrup, R.J.; Cohn, D.H.; Cohen, T.; Byers, P.H.
J. Biol. Chem. 263, 7774-7740, 1988	J. Biol. Chem. 263, 7774-7740, 1988
A;Title: Arginine for glycine substitution in the triple-helical domain of the products of reference number: 155285; NUID:88227975	A;Title: Arginine for glycine substitution in the triple-helical domain of the products of reference number: 155285; NUID:88227975
A;Status: translated from GB/EMBL/DDBJ	A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA	A;Molecule type: DNA
A;Residues: 1090-1107 <WEN1>	A;Residues: 1090-1107 <WEN1>
A;Cross-references: GB:M22816; NID:9179502; PIDN:AAA51844_1; PID:9179503	A;Cross-references: GB:M22816; NID:9179502; PIDN:AAA51844_1; PID:9179503
A;Accession: I70059	A;Accession: I70059
A;Status: translated from GB/EMBL/DDBJ	A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA	A;Molecule type: DNA
A;Residues: 1090-1101, 'R', 1103-1107 <WEN2>	A;Residues: 1090-1101, 'R', 1103-1107 <WEN2>
A;Cross-references: GB:M22817; NID:919606; PIDN:AAA51846_1; PID:9179607	A;Cross-references: GB:M22817; NID:919606; PIDN:AAA51846_1; PID:9179607
A;Note: mutant sequence from a patient with osteogenesis imperfecta type IV	A;Note: mutant sequence from a patient with osteogenesis imperfecta type IV
R;Myers, J.C.; Dickson, L.A.; de Wet, W.J.; Bernard, M.P.; Chu, M.L.; di Liberto, M.; Pe J. Biol. Chem. 258, 1028-1035, 1983	R;Myers, J.C.; Dickson, L.A.; de Wet, W.J.; Bernard, M.P.; Chu, M.L.; di Liberto, M.; Pe J. Biol. Chem. 258, 1028-1035, 1983
A;Title: Analysis of the 3' end of the human pro-alpha-2(I) collagen gene. Utilization of molecular cloning	A;Title: Analysis of the 3' end of the human pro-alpha-2(I) collagen gene. Utilization of molecular cloning
A;Reference number: S09175; NUID:83290853	A;Reference number: S09175; NUID:83290853
A;Accession: S09175	A;Accession: S09175
A;Molecule type: RNA	A;Molecule type: RNA
RESULT 5	A54449
N;Alternative names: procollagen alpha 1(VII) chain precursor - human	N;Alternative names: procollagen alpha 1(VII) chain precursor - human
C;Species: Homo sapiens (man)	C;Species: Homo sapiens (man)
C;Database: 04-Nov-1994 #sequence_revision 04-Nov-1994 #text_change 20-Sep-1999	C;Database: 04-Nov-1994 #sequence_revision 04-Nov-1994 #text_change 20-Sep-1999
C;Accession: A54849; PHO844; S16316; 156328; A30396; 184686	C;Accession: A54849; PHO844; S16316; 156328; A30396; 184686
R;Christiano, A.M.; Greenspan, D.S.; Lee, S.; Ultto, J.	R;Christiano, A.M.; Greenspan, D.S.; Lee, S.; Ultto, J.
J. Biol. Chem. 266, 20256-20262, 1991	J. Biol. Chem. 266, 20256-20262, 1991
A;Title: Cloning of human type VII collagen. Complete primary sequence of the alpha1(I) chain	A;Title: Cloning of human type VII collagen. Complete primary sequence of the alpha1(I) chain
A;Reference number: A54849; NUID:94327588	A;Reference number: A54849; NUID:94327588
A;Accession: A54849	A;Accession: A54849
A;Status: not compared with conceptual translation	A;Status: not compared with conceptual translation
A;Molecule type: mRNA	A;Molecule type: mRNA
A;Residues: 1-2944 <CHR>	A;Residues: 1-2944 <CHR>
A;Cross-references: GB:J02870; NID:9987124; PIDN:AAA75438_1; PID:9987125	A;Cross-references: GB:J02870; NID:9987124; PIDN:AAA75438_1; PID:9987125
R;Tanaka, T.; Takahashi, K.; Furukawa, F.; Immamura, S.	R;Tanaka, T.; Takahashi, K.; Furukawa, F.; Immamura, S.
Biochem. Biophys. Res. Commun. 183, 958-963, 1992	Biochem. Biophys. Res. Commun. 183, 958-963, 1992
A;Title: Molecular cloning and characterization of type VII collagen cDNA.	A;Title: Molecular cloning and characterization of type VII collagen cDNA.
A;Reference number: PHO844; NUID:92231902	A;Reference number: PHO844; NUID:92231902
A;Accession: PHO844	A;Accession: PHO844
A;Molecule type: mRNA	A;Molecule type: mRNA
A;Cross-references: DBP:D1152; DDBJ:D1364; NID:BA02853_1; PID:9536	A;Cross-references: DBP:D1152; DDBJ:D1364; NID:BA02853_1; PID:9536

Proc. Natl. Acad. Sci. U.S.A., 88, 6931-6935, 1991
A: Title: Human type VII collagen: cDNA cloning and chromosomal mapping of the gene.
A: Accession: S16316; MUID:91334380
A: Molecule type: mRNA
A: Residues: 815-892; 'E', '894-1439 <PAR>
A: Cross-references: GB:W65158; GB:S49017; NID:6180914; PIDN:AAA96439_1; PID:910915
A: Experimental source: Keratinocyte
R: Gammon, W.R.; Abernethy, M.L.; Padilla, K.M.; Prisayanh, P.S.; Cook, M.E.; Wright, J.; Invest. Dermatol. 99, 691-696, 1992
A: Title: Noncollagenous (NC1) domain of collagen VII resembles multidomain adhesion protein
A: Reference number: 156328; MUID:93107742
A: Status: translated from GB/EMBL/DBJ
A: Molecule type: mRNA
A: Residues: 'EPR', '372-517', 'D', '520-540', 'W', '542-1255 <RES>
A: Cross-references: GB:S51236; NID:9202308; PIDN:AAB24637_1; PID:9262309
R: Seltzer, J.L.; Eisen, A.Z.; Bauer, E.A.; Morris, N.P.; Granville, R.W.; Burgesson, R.E. J. Biol. Chem. 264, 3822-3826, 1989
A: Title: Cleavage of type VII collagen by interstitial collagenase and type IV collagenase
A: Reference number: A30296; MUID:89139437
A: Accession: A30296
A: Molecule type: protein
A: Residues: 'A', '1240-1246', 'G', '1248-1250', 'X', '1253-1255', 'Q', '1257', 'E', '2032', 'C', '2034-2041'; 'R', 'Greenspan', D.S.
R: Hum. Mol. Genet. 2, 273-278, 1993
A: Title: The carboxyl-terminal half of type VII collagen, including the non-collagenous
A: Reference number: 148103; MUID:93271985
A: Accession: I18486
A: Status: preliminary; translated from GB/EMBL/DBJ
A: Molecule type: mRNA
A: Residues: 2395-2871; 'S', '2873-2944 <RRB2>
A: Cross-references: GB:I06662; NID:9388713; PIDN:AAA89196_1; PID:9388714
R: Christiano, A.M.; Rynnaen, M.; Utton, J. Proc. Natl. Acad. Sci. U.S.A. 91, 3549-3553, 1994
A: Title: Dominant dystrophic epidermolysis bullosa: identification of a Gly --> Ser substitution
A: Reference number: A52525; MUID:94224777
A: Contents: annotation
C: Comment: Prolines and lysines at the third position of the tripeptide repeating unit
 and subsequently O-glycosylated.
C: Genetics:
A: Gene: GDB:COL7A1; EBR1; EBD1; BB
A: Cross-references: GDB:128750; OMIM:120120
A: Map position: 3p21.3-3p21.3
A: Note: defects in this gene can result in dominant and recessive dystrophic epidermolysis bullosa.
A: Note: there are 118 introns
C: Complex: type VII collagen is probably a homotrimer
C: Function:
C: Description: structural component of extracellular polymer associated with anchoring
 f
C: Superfamily: unassigned collagens; animal Kunzitz-type proteinase inhibitor homology; f
C: Keywords: coiled coil; extracellular matrix; glycoprotein; hydroxylysine; hydroxyproline
F: 1-16/Domain: signal sequence #status predicted <SGIC>
F: 17-2944/Product: collagen alpha 1(VII) chain #status predicted <MAT>
F: 17-1253/Domain: amino-terminal nonhelical #status predicted <NC1>
F: 17-1253/Domain: von Willebrand factor type A repeat homology <VWA1>
F: 231-318/Domain: fibronectin type III repeat homology <FN1>
F: 327-413/Domain: fibronectin type III repeat homology <FN2>
F: 414-502/Domain: fibronectin type III repeat homology <FN3>
F: 508-593/Domain: fibronectin type III repeat homology <FN4>
F: 598-683/Domain: fibronectin type III repeat homology <FN5>
F: 686-771/Domain: fibronectin type III repeat homology <FN6>
F: 76-862/Domain: fibronectin type III repeat homology <FN7>
F: 864-952/Domain: fibronectin type III repeat homology <FN8>
F: 954-1045/Domain: fibronectin type III repeat homology <FN9>
F: 1052-1219/Domain: von Willebrand factor type A repeat homology <FN9>
F: 1170-1172/Region: cell attachment (R-G-D) motif
F: 1189-1253/Region: cysteine/proline-rich
F: 1254-2783/Region: interrupted helical
F: 1334-1336/Region: cell attachment (R-G-D) motif
F: 2008-2010/Region: cell attachment (R-G-D) motif
F: 2553-2555/Region: cell attachment (R-G-D) motif

```

F;274-2944/#Domain: carboxyl-terminal nonhelical #status predicted <NCB>
F;2875-2229/#Domain: animal Kinase type proteinase inhibitor homology <BPI>
F;337-785-1109/#Binding site: carboxyl-terminal nonhelical #status predicted
F;2167-2176-2185-2188-2664-2667-2673/#Modified site: 4-hydroxyproline (Pro) #status ex
F;2625-2631/#Modified site: 5-hydroxylysine (Lys) #status experimental
F;2625-2631/#Binding site: carbohydrate (Lys) (covalent) #status experimental
F;2634-2802-2804/#Disulfide bonds: interchain #status predicted

Query Match 65.2%; Score 60; DB 2; Length 2944;
Best Local Similarity 66.7%; Pred. No. 2.2; Mismatches 2; Indels 0; Gaps 0;
Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 1 PAGPMWPNKGKCKVG 15
Db 1623 PPGPVGPRGRIDGEVG 1637

RESULT 6
S10847
collagen alpha 2(I) chain - chicken (fragment)
C;Species: Gallus gallus (chicken)
C;Date: 21-Nov-1993 #sequence_revision 26-Jul-1996 #text_change 23-May-1997
C;Accession: S10847; S65690
R;Dickson, L.A.; Nimonjaya, Y.; Bernard, M.P.; Pesciotta, D.M.; Parsons, J.; Green, G.
J. Biol. Chem. 256: 8407-8415, 1981
A;Title: the exon/intron structure of the 3'-region of the pro-alpha-2(I) collagen gene
A;Reference number: S10847; MUID:81264246
A;Accession: S10847
A;Molecule type: DNA
A;Residues: 1-305 <DIX>
A;Cross-references: EMBL:J00811
A;Note: the authors translated the codon ATA for residue 207 as ASP, AGC for residue
A;Accession: S65690
A;Molecule type: protein
A;Residues: 121-123, 'X', 125-134, 'X', 136-137, 'X', 139, 'XXX', 143 <DIX>
C;Genetics:
A;Gene: COL1A2
A;Introns: 36/3; 54/3; 90/3; 173/1; 236/3; 317/3
C;Superfamily: collagen alpha 2(I) chain; fibrillar collagen carboxyl-terminal homology
C;Keywords: coiled coil; extracellular matrix; glycoprotein; triple helix
F;136-365/#Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match 62.0%; Score 57; DB 2; Length 365;
Best Local Similarity 66.7%; Pred. No. 0.74; Mismatches 2; Indels 0; Gaps 0;
Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 1 PAGPMWPNKGKCKVG 15
Db 56 PHGPGSPHKGKRGNG 70

RESULT 7
A35763
collagen alpha 2 chain - sea urchin (Paracentrotus lividus) (fragment)
C;Species: Paracentrotus lividus (common urchin)
C;Date: 12-Oct-1990 #sequence_revision 12-Oct-1990 #text_change 20-Sep-1999
C;Accession: A35763
R;D'Alessio, M.; Ramirez, F.; Suzuki, H.R.; Solursh, M.; Gambino, R.
J. Biol. Chem. 265: 7050-7054, 1990
A;Title: Cloning of a fibrillar collagen gene expressed in the mesenchymal cells of the
A;Reference number: A35763; MUID:90216744
A;Accession: A35763
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1051 <DAA>
A;Cross-references: GB:J05422; NID:9159961; PIDN:AAA29440.1; PID:9159962
C;Superfamily: unassigned collagens; fibrillar collagen carboxyl-terminal homology <FCC>
F;830-1051/#Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match 62.0%; Score 57; DB 2; Length 1051;
Best Local Similarity 66.7%; Pred. No. 2.2; Mismatches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

```

Qy 1 PAGGWWGPNCNDGKVG 15
 ||||| | | | : | : |
 C;Species: *Mus musculus* (house mouse)
 C;DBatt: 28-May-1992 #sequence_revision 28-May-1992 #text_change 13-Aug-1999
 C;Accession: A41182; A44885
 R;Metsaeranta, M.; Toman, D.; de Crombrugge, B.; Vuorio, E.
 J. Biol. Chem. 266, 16862-16869, 1991
 A;Title: Mouse type II collagen gene. Complete nucleotide sequence, exon structure, and
 accession number: A41182; MUID:91358489
 A;Status: preliminary; not compared with conceptual translation
 A;Molecule type: DNA
 A;Residues: 1-149 <MET>
 A;Cross-references: GB:M65161
 R;Cheah, K.S.; Lau, E.T.; Au, P.K.; Tam, P.P.
 development 111, 945-953, 1991
 A;Title: Expression of the mouse alpha 1(II) collagen gene is not restricted to cartilage
 A;Accession: A44885; MUID:91347939
 A;Molecule type: DNA
 A;Residues: 1-28 <CHE>
 A;Cross-references: GB:S63190; NID:9234368; PIDN:AAB19627.1; PID:9234369
 A;Note: sequence extracted from NCBI backbone (NCBIN63190, NCBIPI:63192)
 C;Keywords: alternative splicing; coiled coil; extracellular matrix; glycoprotein; trimer
 F;1191-1419/Domain: fibrillar collagen carboxyl-terminal homology <FC>
 Query Match 62.0%; Score 57; DB 2; Length 1419;
 Best Local Similarity 73.3%; Pred. No. 2.9;
 Matches 11; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 Qy 1 PAGGWWGPNCNDGKVG 15
 ||||| | | | : | : |
 Db 848 PAGGPGPAGKDGPKG 862
 RESULT 9
 B41182
 Collagen alpha 1(II) chain precursor (long splice form) - mouse
 C;Species: *Mus musculus* (house mouse)
 C;DBatt: 28-May-1992 #sequence_revision 28-May-1992 #text_change 16-Jul-1999
 C;Accession: B41182
 R;Metsaeranta, M.; Toman, D.; de Crombrugge, B.; Vuorio, E.
 J. Biol. Chem. 266, 16862-16869, 1991
 A;Title: Mouse type II collagen gene. Complete nucleotide sequence, exon structure, and
 accession number: A41182; MUID:91358489
 A;Status: preliminary; not compared with conceptual translation
 A;Molecule type: DNA
 A;Residues: 1-1487 <MET>
 A;Cross-references: GB:M65161
 C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
 C;Keywords: alternative splicing; coiled coil; extracellular matrix; glycoprotein; trimer
 F;1259-1487/Domain: fibrillar collagen carboxyl-terminal homology <FC>
 Query Match 62.0%; Score 57; DB 2; Length 1487;
 Best Local Similarity 73.3%; Pred. No. 3.1;
 Matches 11; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 Qy 1 PAGGWWGPNCNDGKVG 15
 ||||| | | | : | : |
 Db 916 PAGGPGPAGKDGPKG 930
 RESULT 10
 S07133

RESULT 8
 A41182
 collagen alpha 1(III) chain precursor - mouse
 C;Species: *Gallus gallus* (chicken)
 C;DBatt: 13-Jan-1995 #sequence_revision 17-Jul-1998 #text_change 13-Aug-1999
 C;Accession: S0133; S2447
 R;Deak, F.; Argraves, W.S.; Kiss, I.; Sparks, K.J.; Goetinck, P.F.
 Biochem. J. 229, 189-196, 1985
 A;Title: Primary structure of the telopeptide and a portion of the helical domain of
 collagen alpha 1(III) chain precursor - chicken (fragment)
 A;Reference number: S07133; MUID:8530662
 A;Accession: S07133
 A;Status: preliminary
 A;Molecule type: mRNA
 A;Residues: 1-193 <DEA>
 A;Cross-references: EMBL:X02663; NID:963318; PIDN:CAN26499.1; PID:9755732
 R;van der Rest, M.; Mayne, R.
 J. Biol. Chem. 263, 1615-1618, 1988
 A;Title: Type IX collagen proteoglycan from cartilage is covalently cross-linked to t
 A;Reference number: S23813; MUID:8811574
 A;Accession: S24447
 A;Molecule type: protein
 A;Residues: 124-128, 'X', 127-129, 'AO' <RES>
 C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homolo
 Query Match 60.9%; Score 56; DB 2; Length 193;
 Best Local Similarity 66.7%; Pred. No. 0.55;
 Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
 Qy 1 PAGGWWGPNCNDGKVG 15
 ||||| | | | : | : |
 Db 47 PPGPVGPGSKDGGSNG 61
 RESULT 9
 T19289
 hypothetical protein C15A11.1 - *caenorhabditis elegans*
 C;Species: *Caenorhabditis elegans*
 C;DBatt: 15-Oct-1999 #sequence_revision 15-oct-1999 #text_change 21-Jan-2000
 C;Accession: T19289
 R;Gardner, A.
 submitted to the EMBL Data Library, September 1996
 A;Reference number: Z19103
 A;Accession: T19289
 A;Gene: C15A11.1
 A;MP position: 1
 A;Molecule type: DNA
 A;Residues: 1-303 <WIL>
 A;Cross-references: EMBL:279694; PIDN:CAB01959.1; GSPPDB:GN00019; CESP:C15A11.1
 A;Experimental source: clone C15A11
 C;Genetics:
 A;Status: preliminary; translated from GB/EMBL/PDB/J
 A;Cross-references: 1-303 <WIL>
 A;Reference number: Z19103
 A;Gene: C15A11.1
 A;MP position: 1
 A;Molecule type: DNA
 A;Residues: 1-303 <WIL>
 C;Superfamily: unassigned collagens
 Query Match 60.9%; Score 56; DB 2; Length 303;
 Best Local Similarity 66.7%; Pred. No. 0.86;
 Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
 Qy 1 PAGGWWGPNCNDGKVG 15
 ||||| | | | : | : |
 Db 210 PPGPGPGPKDGEG 224
 RESULT 10
 S41067
 collagen alpha 1(III) chain - rat
 C;Species: *Rattus norvegicus* (Norway rat)
 C;DBatt: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 13-Aug-1999
 C;Accession: S41067; S23905; S31924
 R;Glumoff, V.; Maekelae, J.K.; Vuorio, E.
 Biochim. Biophys. Acta 1217, 41-48, 1994
 A;Title: Cloning of cDNA for rat pro alpha 1(III) collagen mRNA. Different expression
 A;Reference number: S41067; MUID:94114571
 A;Accession: S41067
 A;Status: preliminary
 A;Molecule type: mRNA

A;Residues: 1-636 <GLUD>
A;Cross-references: EMBL:X70369; NID:957915; PIDN:CAA49832.1; PID:g57916
R;Frankel, F.R.; Hsu, C.Y.J.; Myers, J.C.; Lin, E.; Lytle, C.R.; Komm, B.; Mohn, K.
DNA 7, 347-354, 1988
A;Title: Regulation of alpha-2 (I)-alpha-1 (III), and alpha-2(V) collagen mRNAs by ester
A;Reference number: A29905; MUID:88296083
A;Accession: A29905
A;Molecule type: mRNA
A;Residues: 308-482 <FRRA>
A;Cross-references: GB:M21354; NID:9203500; PIDN:AAA40942.1; PID:g203501
R;Glunoff, V.; Maekelae, J.K.; Vuorio, E.
R;Submitted to the EMBL Data Library, February 1993
A;Status: preliminary
A;Accession: S31924
A;Accession: S31924
A;Molecule type: mRNA
A;Residues: 2-636 <GL2>
A;Cross-references: EMBL:X70369
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
C;Keywords: coiled coil; extracellular matrix; glycoprotein; trimer; triple helix
F;408-636/Domain: fibrillar collagen carboxyl-terminal homology <FCC>
Query Match Score 56; DB 2; Length 1076;
Best Local Similarity 66.7%; Pred. No. 3.1; Mismatches 4; Indels 0; Gaps 0;
Matches 10; Conservative 1; MisMatches 1;
Qy 1 PAGPWNPGNGKDKGVG 15
Db 313 PVGPHGPKGSG 327

RESULT 13
S28791
collagen alpha 1(XI) chain - chicken (fragment)
C;Species: Gallus gallus (chicken)
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 13-Aug-1999
C;Accession: S28791
R;Nah, H.D.; Barembaum, M.; Upholt, W.B.
J. Biol. Chem. 267, 22581-22586, 1992
A;Title: The chicken alpha1(XI) collagen gene is widely expressed in embryonic tissues.
A;Reference number: S28791; MUID:93054557
A;Accession: S28791
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-688 <RNA>
A;Cross-references: EMBL:N88593; NID:g211619; PIDN:AAA48707.1; PID:g211620
C;Superfamily: collagen alpha 1(V) chain; fibrillar collagen carboxyl-terminal homology <FCC>
F;665-887/Domain: fibrillar collagen carboxyl-terminal homology <FCC>
Query Match Score 56; DB 2; Length 636;
Best Local Similarity 66.7%; Pred. No. 1.8; Mismatches 5; Indels 0; Gaps 0;
Matches 10; Conservative 0; MisMatches 5;
Qy 1 PAGPWNPGNGKDKGVG 15
Db 313 PVGPHGPKGSG 327

RESULT 14
T30842
serine-repeat antigen 3 - Plasmodium vivax
C;Species: Plasmodium vivax
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 11-May-2000
C;Accession: T30842
R;Kisler, M.C.; Crawford, K.A.; Boley, L.J.; Landsberg, K.E.; Gibson, H.L.; Kaslow, D.C.
Mol. Biochem. Parasitol. 78, 55-65, 1996
A;Title: Identification and cloning of a locus of serine repeat antigen (sera)-related genes
A;Reference number: 220898; MUID:86408670
A;Accession: T30842
A;Status: preliminary; translated from GB/EMBL/DDJB
A;Molecule type: DNA
A;Residues: 1-1076 <KIE>
A;Cross-references: EMBL:051723; NID:g1381087; PID:g1381090; PIDN:AAB41487.1

RESULT 15
B40333
collagen alpha 1(II) chain precursor - African clawed frog
C;Species: Xenopus laevis (African clawed frog)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C;Accession: B40333
R;Su, M.W.; Suzuki, H.R.; Bieler, J.J.; Solursh, M.; Ramirez, F.
J. Cell Biol. 115, 565-575, 1991
A;Title: Expression of two nonallelic type II procollagen genes during Xenopus laevis
A;Reference number: A40333; MUID:92011838
A;Accession: B40333
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1486 <UUA>
A;Cross-references: GB:M3595
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology <FCC>
F;1258-1486/Domain: fibrillar collagen carboxyl-terminal homology <FCC>
Query Match Score 56; DB 1; Length 1486;
Best Local Similarity 66.7%; Pred. No. 4.3; Mismatches 4; Indels 0; Gaps 0;
Matches 10; Conservative 1; MisMatches 4;
Qy 1 PAGPWNPGNGKDKGVG 15
Db 1167 PPGPVGPGSKGGSNG 1181

RESULT 16
A40333
collagen alpha 1'(III) chain precursor - African clawed frog
C;Species: Xenopus laevis (African clawed frog)
C;Date: 16-Sep-1992 #sequence_revision 16-Sep-1992 #text_change 16-Jul-1999
C;Accession: A40333
R;Su, M.W.; Suzuki, H.R.; Bieker, J.J.; Solursh, M.; Ramirez, F.
J. Cell Biol. 115, 565-575, 1991
A;Title: Expression of two nonallelic type III procollagen genes during Xenopus laevis
A;Reference number: A40333; MUID:92011838
A;Accession: A40333
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-1492 <SUAA>
A;Cross-references: GB:M3596
A;Note: this sequence is presented as substitutions relative to another sequence in a
es they replace; the appropriate interpretation of the sequence figure was reconstruc
C;Superfamily: collagen alpha 1(II) chain; fibrillar collagen carboxyl-terminal homolo
C;Keywords: coiled coil; extracellular matrix; glycoprotein; trimer; triple helix
F;37-96/Domain: von Willebrand factor type C repeat homology <WCC>
F;1263-1492/Domain: fibrillar collagen carboxyl-terminal homology <FCC>
Query Match Score 56; DB 2; Length 1492;
Best Local Similarity 66.7%; Pred. No. 4.3; Mismatches 4; Indels 0; Gaps 0;
Matches 10; Conservative 1; MisMatches 4;
Qy 1 PAGPWNPGNGKDKGVG 15
Db 1170 PPGPVGPGSKGGSNG 1184

RESULT 17

CGHUE
Collagen alpha 1(XI) chain precursor - human
N; Alternate names: procollagen alpha 1(XI) chain
C; Date: 31-Mar-1990 #sequence_revision 03-Oct-1995 #text_change 08-May-1998
C; Accession: A35239; A3195
R; Yoshioka, H.; Ramirez, F.
J. Biol. Chem. 265, 6423-6426, 1990
A; Title: Pro-alpha1(XI) collagen. Structure of the amino-terminal propeptide and expression
A; Reference number: A35239; MUID:90202924
A; Accession: A35239
A; Molecular type: mRNA
A; Residues: 1-558 <XOS>
A; Cross-references: GB:005407
A; Reference number: A92689; MUID:89034222
A; Accession: A31795
A; Molecular type: DNA; mRNA
A; Residues: 530-1806 <BERZ>
A; Cross-references: GB:004177
A; Note: parts of this sequence were determined by protein sequencing
C; Comment: Prolines and lysines at the third position of the tripeptide :repeating unit
ed and subsequently O-glycosylated.

C; Genetics:

A; Gene: GDB:COL1A1; COL6

A; Cross-references: GDB:120555; OMIM:120280

A; Map position: 1p21-1p21

A; Introns: 561/3; 579/3; 597/3; 615/3; 633/3; 648/3; 666/3; 681/3

A; Note: the list of introns is incomplete

C; Complex: type XI collagen may be a heterotrimer of two alpha 1(XI) chains and one alpha 3(XI) chain (see PIR:CGHUC), initially linked by disulfide bonds among their carboxyl-terminal with desmosine cross-links made from lysine and allysine residues

C; Function:

A; Description: structural component of extracellular fibrous polymer associated with cell

A; Note: may play a role in controlling the lateral growth of collagen II fibrils

C; Superfamily: collagen alpha 1(V) chain; fibrillar collagen carboxyl-terminal homology

C; Keywords: coiled coil; extracellular matrix; glycoprotein; hydroxylysine; hydroxyproline; 1-36/Domain: signal sequence #status predicted <SIG>

F; 1-26/Domain: PARR-like #status predicted <PAR>

F; 37-51/Domain: amino-terminal propeptide #status predicted <PRO>

F; 512-1567/Product: collagen alpha 1(XI) chain #status predicted <MAT>

F; 512-57/Region: amino-terminal nonhelical telopeptide

F; 1543-1565/Region: carboxy-terminal nonhelical telopeptide

F; 1566-1806/Domain: carboxyl-terminal propeptide #status predicted <CTP>

F; 1583-1805/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

F; 161-243-182-236/Disulfide bonds: #status predicted

F; 505/Modified site: allysine (Lys) #status predicted

F; 612-1457/Modified site: 5-hydroxylysine (Lys) #status predicted

F; 612,1452/Binding site: carbohydrate (Lys) (covalent) #status predicted

Query Match 60.9%; Score 56; DB 1; Length 1806; Best Local Similarity 66.7%; Pred. No. 5-2; Matches 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0; Qy 1 PAGPMGPNGDKYVG 15 Db 1226 PRGPGPQGADGPOG 1240

RESULT 18

D41132
collagen-related protein 4 - Hydra magnipapillata (fragment)
C; Species: Hydra magnipapillata
C; Date: 05-Jun-1992 #sequence_revision 12-Jun-1992 #text_change 01-Dec-2000
C; Accession: D41132; S21932
R; Kurz, E.M.; Holstein, T.W.; Petri, B.M.; Engel, J.; David, C.N.
J. Cell Biol. 115, 1159-1169, 1991

Query Match 59.8%; Score 55; DB 2; length 363; Best Local Similarity 66.7%; Pred. No. 1.4; Matches 10; Conservative 1; Mismatches 1; Gaps 0; Qy 1 PAGPMGPNGDKYVG 15 Db 167 PAGPRGPNGNSGSDG 181

RESULT 19

T16841
hypothetical protein T10E10.2 - Caenorhabditis elegans
C; Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 02-Jun-2000
C; Accession: T16841
R; Geisel, C.
Submitted to the EMBL Data Library, October 1995
A; Description: The sequence of C. elegans cosmid T10E10.
A; Gene: Z1888
A; Accession: T16841
A; Status: preliminary; translated from GB/EMBL/DDJB
A; Residues: 1-326 <GEI>
A; Cross-references: EMBL:U39644; NID:91049339; PID:91049341; PIDN:AAA80350.1; CESP:T1
C; Genetics:
A; Gene: CESP:T10E10.2
C; Superfamily: unassigned collagens
Query Match 59.8%; Score 55; DB 2; Length 326; Best Local Similarity 66.7%; Pred. No. 1.3; Matches 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0; Qy 1 PAGPMGPNGDKYVG 15 Db 167 PAGPRGPNGNSGSDG 181

RESULT 20

T16831
hypothetical protein T07H6.3 - Caenorhabditis elegans
C; Species: Caenorhabditis elegans
C; Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 18-Feb-2000
C; Accession: T16831
R; Geisel, C.
Submitted to the EMBL Data Library, April 1996
A; Description: The sequence of C. elegans cosmid T07H5.
A; Reference number: 218566
A; Accession: T16831
A; Status: preliminary; translated from GB/EMBL/DDJB
A; Molecular type: DNA
A; Residues: 1-363 <GEI>
A; Cross-references: EMBL:U53344; NID:91255886; PID:91255887; PIDN:AAA96223.1; GSPDB:G
A; Experimental source: Strain Bristol N2; clone T07H6
C; Genetics:
A; Gene: CESP:T07H6.3
A; Map position: X
A; Introns: 62/3
C; Superfamily: unassigned collagens
Query Match 59.8%; Score 55; DB 2; length 363; Best Local Similarity 66.7%; Pred. No. 1.4; Matches 10; Conservative 1; Mismatches 1; Gaps 0; Qy 1 PAGPMGPNGDKYVG 15 Db 167 PAGPRGPNGNSGSDG 181

Matches 10; conservative 0; Mismatches 5; Indels 0; Gaps 0;
Qy 1 PAGPWPNGKDGKVG 15
Db 229 PAGPKGPNGNSGSDG 243

RESULT 21

559513 collagen II A1 protein - zebra fish (fragment)
C;Species: Brachydanio rerio (zebra fish)
C;Date: 12-Feb-1998 #sequence_revision 20-Feb-1998 #text_change 13-Aug-1999
C;Accession: S59513
A;Status: preliminary
R;Yan, Y.; Riddleman, B.; Postlethwait, J.H.
A;Submitted to the EMBL Data Library, March 1995
A;Description: Expression of a type II collagen gene in the zebrafish embryonic axis.
A;Accession: S59513
A;Molecule type: mRNA
A;Residues: 1-164 <YAM>
A;Cross-references: EMBL:U23822; NID:973660; PIDN:AAA96815.1; PID:973661
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology; F;236-464>/domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match 59.8%; Score 55; DB 2; Length 615;
Best Local Similarity 66.7%; Pred. No. 1.9;
Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
Qy 1 PAGPWPNGKDGKVG 15
Db 142 PSGPVPGPAGKGSGNG 156

RESULT 22

A05269 collagen alpha 1(II) chain precursor - chicken (fragments)
C;Species: Gallus gallus (chicken)
C;Date: 05-Jun-1987 #sequence_revision 05-Jun-1987 #text_change 16-Jul-1999
C;Accession: A05269; A38035; A20855
R;Yamada, Y.; Liang, G.; Mudryj, M.; Obici, S.; de Crombrughe, B.
Nature 310, 333-337, 1984
A;Title: Conservation of the sizes for one but not another class of exons in two chick c
A;Reference number: A05269; MUID:8427056
A;Accession: A05269
A;Molecule type: DNA
A;Residues: 1-129 <YAM1>
A;Note: the authors translated the codon AAA for residue 68 as Arg, AAC for residue 173
R;Yamada, Y.; Ruehn, K.; de Crombrughe, B.
Nucleic Acids Res. 11, 2733-2744, 1983
A;Reference number: A38035; MUID:83220816
A;Accession: A38035
A;Molecule type: DNA
A;Residues: 330-615 <YAM2>
A;Note: the authors translated the codon GAT for residue 548 as Glu
R;Yamada, Y.; Mudryj, M.; Sullivan, M.; de Crombrughe, B.
J. Biol. Chem. 258, 2758-2761, 1983
A;Title: Isolation and characterization of a genomic clone encoding chick alpha1 type II
A;Reference number: A20855; MUID:83135706
A;Accession: A20855
A;Molecule type: DNA
A;Residues: 222-233, 'R', 235-236, 'V', 238-239/330-336, 'V', 338-351 <YAM>
A;Note: the authors translated the codon CGT for residue 234 as GLY, GTT for residue 237
C;Comment: Chicken collagen alpha 1(III) chain has about 50 exons. This sequence correspond
9/240-275, 6/276-293), 5/294-329), 4/(330-423), 3(424-484), 2(485-567), and 1(568-615)
C;Genetics: Introns: 68/3; 86/3; 119/3; 134/3; 152/3; 188/3; 221/3; 239/3; 275/3; 293/3; 423/3; 46
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology; F;2-236-464>/domain: amino-terminal propeptide nonhelical
F;2-68/Region: amino-terminal propeptide nonhelical
F;2-68/Domain: von Willebrand factor type C repeat homology <YWC>
F;69-347>/domain: collagen alpha helical chain (fragments) #status predicted <CAH>
F;348-615>/domain: carboxyl-terminal propeptide #status predicted <CTP>

RESULT 23

CGCH2S collagen alpha 2(I) chain Precursor - chicken (fragments)
C;Species: Gallus gallus (chicken)
C;Date: 24-Apr-1984 #sequence_revision 15-Aug-1997 #text_change 21-Jul-2000
C;Accession: I50206; I50207; S07354; S10848; S10400; S11146; I50628; I50170; I50625;
R;Boedtker, H.; Finer, M.; Ahn, S.; Ann, N. Y. Acad. Sci. 460, 85-116, 1985
A;Title: The structure of the chicken alpha 2 collagen gene.
A;Reference number: I50206; MUID:86185168
A;Accession: I50206
A;Status: translated from GB/EMBL/DDJB
A;Molecule type: DNA
A;Residues: 1-245 <YAM>
A;Cross-references: GB:M25963; NID:9211581; PIDN:AAA69960.1; PID:g211605
A;Accession: I50207
A;Status: translated from GB/EMBL/DDJB
A;Molecule type: DNA
A;Residues: 246-431 <BO2>
A;Cross-references: GB:M25965; NID:9211583; PIDN:CAA25330.1; PID:9211606
R;Aho, S.; Tate, V.; Boedtker, H.
Nucleic Acids Res. 12, 6117-6125, 1984
A;Title: Location of the 11 bp exon in the chicken pro alpha-2(I) collagen gene.
A;Reference number: S07354; MUID:84297217
A;Accession: S07354
A;Molecule type: DNA
A;Cross-references: EMBL:X00760; NID:963266; PIDN:CAA25330.1; PID:963267
R;Rate, V.E.; Finer, M.H.; Boedtker, H.; Doty, P.
Nucleic Acids Res. 11, 91-104, 1983
A;Title: Chick pro-alpha-2(I) collagen gene: exon location and coding potential for
A;Reference number: S10480; MUID:83246518
A;Accession: S10480
A;Molecule type: mRNA
A;Residues: 1-89 <YAM>
A;Cross-references: EMBL:X02657; NID:963314; PIDN:CAA26493.1; PID:963315
A;Accession: S10480
A;Molecule type: DNA
A;Residues: 17-73 <YAM>
A;Note: the authors translated the codon CAG for residue 42 as Glu
R;Vogeli, G.; Ohkubo, H.; Sobel, M.E.; Yamada, Y.; Pastan, I.; de Crombrughe, B.
Proc. Natl. Acad. Sci. U.S.A. 78, 5334-5338, 1981
A;Title: Structure of the promoter for chicken alpha-2 type I collagen gene.
A;Reference number: S11146; MUID:82060240
A;Accession: S11146
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-15 <YAM>
A;Cross-references: EMBL:J00821
R;Yamada, Y.; Avvedimento, V.E.; Mudryj, M.; Ohkubo, H.; Vogeli, G.; Irani, M.; Pastan
Cell 22, 887-892, 1980
A;Title: The collagen gene: evidence for its evolutionary assembly by amplification of
A;Reference number: I50170; MUID:8112157
A;Accession: I50170
A;Status: preliminary; translated from GB/EMBL/DDJB
A;Molecule type: DNA
A;Residues: 126-161 <YAM>
A;Cross-references: EMBL:V00400; NID:963305; PID:9833611
A;Accession: I50170
A;Status: preliminary; translated from GB/EMBL/DDJB

A; Molecule type: DNA
A; Residues: 126-161 <YA2>
A; Cross-references: GB:J00828; NID:9211295; PIDN:AAA51612.1.; PID:9211317
A; Accession: I50625
A; Status: preliminary; translated from GB/EMBL/DDBJ
A; Molecule type: DNA
A; Residues: 530-558 <YA3>
A; Cross-references: EMBL:V00396; NID:963295; PID:9833609
A; Accession: I50626
A; Status: preliminary; translated from GB/EMBL/DDBJ
A; Molecule type: DNA
A; Residues: 572-601 <YA4>
A; Cross-references: EMBL:V00398; NID:963299; PID:9833610
A; Accession: I50624
A; Status: preliminary; translated from GB/EMBL/DDBJ
A; Molecule type: DNA
A; Residues: 433-443 <YA5>
A; Cross-references: EMBL:V00394; NID:963287; PID:9833608
R; Kang, A.H., Gross, J.
Biochemistry 9, 796-804, 1970
A; Title: Amino acid sequence of cyanogen bromide peptides from the amino-terminal region
A; Reference number: A90568; MUID:70131186
A; Accession: A90568
A; Molecule type: protein
A; Residues: 72 ,79-92 <KAN>
A; Experimental source: skin
R; Kang, A.H.; Igashii, S.; Gross, J.
Biochemistry 8, 3200-3204, 1969
A; Title: Characterization of the cyanogen bromide peptides from the alpha₂ chain of chick
A; Reference number: A90557; MUID:69205369
A; Accession: A90557
A; Molecule type: protein
A; Residues: 2 ,79-95;398-409, 'A', 411, 'V', 413-428 <KA2>
A; Experimental source: skin
A; Note: the compositions of the six CNBr peptides were determined. CNBr0 is residues 93-
R; Higlberger, J.H.; Kang, A.H.; Gross, J.
Biochemistry 10, 610-616, 1971
A; Title: Comparative studies on the amino acid sequence of the alpha₂-CB2 peptides from
A; Reference number: A90358; MUID:71115216
A; Accession: B90358
A; Molecule type: protein
A; Residues: 399-409, 'A', 411, 'V', 413-428 <HIG>
A; Experimental source: skin
R; Lane, J.M.; Miller, E.J.
Biochemistry 8, 2134-2139, 1969
A; Title: Isolation and characterization of the peptides derived from the alpha₂ chain of
A; Reference number: A90555; MUID:69206882
A; Accession: A90555
A; Molecule type: protein
A; Residues: 'Z' ,79-95;398-409, 'A', 411, 'V', 413-428 <LAN>
A; Note: the compositions of the six CNBr peptides were determined
A; Note: the compositions of CNBr1, CNBr0, and CNBr2 are identical with those from skin a
R; Igashii, S.; Kang, A.H.; Gross, J.
Biochem. Biophys. Res. Commun. 38, 697-702, 1970
A; Title: Renaturation and ordering by electron microscopy of the cyanogen bromide peptide
A; Reference number: A90168; MUID:70181851
A; Contents: annotation; skin; order of CNBr peptides
R; Vuust, J.; Lane, J.M.; Fietzke, P.P.; Miller, E.J.; Piez, K.A.
Biochem. Biophys. Res. Commun. 38, 703-708, 1970
A; Title: The order of the CNBr peptides from the alpha₂ chain of collagen.
A; Reference number: A90169; MUID:70181852
A; Content: annotation; bone; order of CNBr peptides
R; Mooney, J.; Hanahan, D.; Tate, V.; Boedtker, H.; Doty, P.
Nature 294, 129-135, 1981
A; Title: Structure of the pro alpha₂(I) collagen gene.
A; Reference number: S0327; MUID:8205081
A; Accession: S07327
A; Molecule type: DNA
A; Residues: 94-92135, 247-431;432,693-774 <MO22>
A; Cross-references: EMBL:J00826
R; Fuller, F.; Boedtker, H.
Biochemistry 20, 996-1006, 1981
A; Title: Sequence determination and analysis of the 3' region of chicken pro-alpha 1 (I)
A; Reference number: 150623; MUID:81160715
A; Accession: I50623
A; Status: preliminary; translated from GB/EMBL/DDBJ
A; Molecule type: mRNA
A; Residues: 506-830, 'T', 831-903, 'N', 904-964 <FUL>
A; Cross-references: EMBL:V00390; NID:963248; PIDN:CAA23688.1.; PID:963249
R; Avendanno, E.V.; Vogeli, G.; Yamada, Y.; Maizel, J.V.
Cell 21, 689-696, 1980
A; Title: Correlation between splicing sites within an intron and their sequence compl
A; Reference number: 150172; MUID:81064671
A; Status: preliminary; translated from GB/EMBL/DDBJ
A; Molecule type: DNA
A; Residues: 536-558 <AVW>
A; Cross-references: GB:MI0581; NID:9211323; PIDN:AAA48637.1.; PID:9211326
R; Lehrach, H.; Frischauft, A.M.; Hanahan, D.; Wooley, J.; Fuller, F.; Crkvenjakov, R.;
Proc. Natl. Acad. Sci. U.S.A. 75, 47-542, 1978
A; Title: Construction and characterization of a 2.5-kilobase procollagen clone.
A; Reference number: 150171; MUID:79074829
A; Accession: I50171
A; Status: preliminary; translated from GB/EMBL/DDBJ
A; Molecule type: mRNA
A; Residues: 484-505 <LEH>
A; Cross-references: GB:J00837; NID:94530617; PIDN:AAA51614.1.; PID:9211320
C; Genetics:
C; Gen: COLA2
A; Introns: 24/1; 27/3; 33/3; 45/3; 73/3; 92/3; 107/3; 125/3; 143/3; 161/3; 179/3; 197/
A; Note: the list of introns is incomplete
C; Superfamily: collagen alpha 2(I) chain; fibrillar collagen carboxyl-terminal homolo
C; Keywords: blocked amino end; coiled coil; extracellular matrix; glycoprotein; hydroo
F; 1-22/Domain: signal sequence #status predicted <SIG>
F; 23-77/Domain: amino-terminal propeptide #status predicted <PRO>
F; 78-964/Product: collagen alpha 2(I) chain (fragments) #status predicted <MATN>
F; 737-894/Region: amino-terminal nonhelical telopeptide
F; 737-964/Domain: fibrillar collagen carboxyl-terminal homology <FCC>
F; 8/Modified site: blocked amino end (Gin) (in mature form) (probably pyrrolidone ca
F; 83/Modified site: allysine (Lys) #status experimental
F; 224-245/Modified site: 4-hydroxyproline (Pro) #status experimental
F; 866/Binding site: carbohydrate (Asn) (covalent) #status predicted
Query Match 59 8%; score 55; DB 1; Length 964;
Best Local Similarity 66.7%; Pred. No. 3.9;
Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
Qy 1 PAGGWGPNGDKGV 15
Db 657 PPGFSGPPGKDGKNG 671
RESULT 24
CGCHS
collagen alpha 1(I) chain - chicken (tentative sequence) (fragments)
C;Species: Gallus gallus (chicken)
C;Date: 12-Aug-1981 #sequence_revision 06-Jul-1982 #text_change 31-Mar-2000
C;Accession: A90458; A90181; A02837
R; Higlberger, J.H.; Cobbett, C.; Dixit, S.N.; Yu, W.; Seyer, J.M.; Kang, A.H.; Gross,
Biochemistry 21, 2048-2055, 1982
A; Title: Amino acid sequence of chick skin collagen alpha1(I)-GB8 and the complete pri
A; Reference number: A90458; MUID:82231995
A; Accession: A90458
A; Molecule type: protein
A; Residues: 1-1036 <HIG>
A; Status: preliminary; translated from GB/EMBL/DDBJ
A; Molecule type: mRNA
A; Residues: 1-1036 <SKIN>
A; Note: this is the latest in a series of papers from these workers elucidating the s
R; Doty, D.R.; Glimcher, M.J.
Biochem. Biophys. Res. Commun. 48, 720-726, 1972
A; Title: Evidence for a previously undetected sequence at the carboxyterminus of the
A; Reference number: A90181
A; Accession: A90181
A; Molecule type: protein
A; Residues: 1037-1042 <YR>
A; Experimental source: skin

A;Note: residues 1037-1042 above correspond to the carboxyl end of the protein
 C;Comment: lysines at positions 103, 700, 934, and 946 above may be hydroxylated in some
 C;Comment: Most of the prolines at the third position of the tripeptide repeating unit
 C;Comment: Pro-1002 is the only 3-hydroxyproline and the only hydroxylated proline in pc
 C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
 C;Keywords: coiled coil; extracellular matrix; glycoprotein; pyroglutamic acid; trimer;
 F;1/2/modified site: pyrrolidone carboxylic acid (Gln) #status experimental

Query Match	59.8%	Score 55;	DB 1;	Length 1042;		
Best Local Similarity	66.7%	Pred. No.	4.2;			
Matches	10;	Conservative	1;	Mismatches		
Qy	1 PAGPWNPGNGDGKV 15	4;	Indels	0;	Gaps	0;
Db	60 PAGPPGKNGDDGEAG 74					

RESULT 25
 T45467
 collagen alpha 1(II) chain precursor [imported] - horse
 N;Alternate names: type II collagen
 C;Species: Equus caballus (domestic horse)
 C;Date: 31-Jan-2000 #sequence_revision 31-Jan-2000
 C;Accession: T45467
 R;Richardson, D.W.; Dodge, G.R.
 R;Submitted to the EMBL Data Library, June 1996
 A;Description: Cloning of equine type II collagen and modulation of its expression in eq
 A;Reference number: 222977
 A;Accession: T45467
 A;Status: preliminary; translated from GB/EMBL/DDJB
 A;Molecule type: mRNA
 A;Residues: 1-1418 <RIC>
 A;Cross-references: EMBL:U62528; PIDN:AAB05773.1
 C;Superfamily: collagen alpha 1(II) chain; fibrillar collagen carboxyl-terminal homology;
 Query Match

Best Local Similarity	59.8%	Score 55;	DB 2;	Length 1464;		
Matches	10;	Conservative	5.7;			
Qy	1 PAGPWNPGNGDGKV 15	4;	Indels	0;	Gaps	0;
Db	1096 PPGPPGPGSKGDGANG 110					

RESULT 26
 S59856
 collagen alpha 1(III) chain precursor - mouse
 C;Species: Mus musculus (house mouse)
 C;Accession: 10-Apr-1996 #sequence_revision 19-Apr-1996 #text_change 13-Aug-1999
 R;Toman, P.D.; de Crombrugge, B.
 Gene 147, 161-168, 1994
 A;Title: The mouse type-III procollagen-encoding gene: genomic cloning and complete DNA
 A;Reference number: S59856; MUID:95011609
 A;Accession: S59856
 A;Molecule type: DNA
 A;Residues: 1-1464 <TOM>
 A;Cross-references: EMBL:X52046
 R;Toman, D.
 submitted to the EMBL Data Library, November 1994
 A;Reference number: S62120
 A;Accession: S62120
 A;Molecule type: DNA
 A;Residues: 1-866; G', 868-1464 <TOA>
 A;Cross-references: EMBL:X52046; MUID:9575321; PIDN:CAA36279.1; PID:9575322
 R;Metsaeranta, M.; Toman, D.; de Crombrugge, B.; Vuorio, E.
 Biochim. Biophys. Acta 1089, 241-243, 1991
 A;Title: Specific probes for mouse type I, II, III and IX collagen mRNAs.
 A;Reference number: S16176; MUID:91274355
 A;Accession: S16373
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1442-1464 <MET>

RESULT 27
 CGHUGC
 collagen alpha 1(III) chain precursor [validated] - human
 N;Alternate names: Procollagen alpha 1(III) chain
 N;Contains: chondrocalcin; collagen alpha 1(III) chain precursor splice form 1; collag
 C;Species: Homo sapiens (man)
 C;Date: 28-May-1986 #sequence_revision 01-Sep-1995 #text_change 08-Dec-2000
 C;Accession: A38513; S06715; S24270; A24828; S06496; A35428; A30147; A33116; S64674;
 T250; I37751; I37252; I37253; I37254; I35338; I59535; I61910
 R;Ryan, M.C.; Sierski, M.; Sandell, L.J.
 Genomics 8, 41-48, 1990
 A;Title: The human type II procollagen gene: identification of an additional protein-
 A;Reference number: A38513; MUID:91184811
 A;Accession: A38513
 A;Molecule type: DNA
 A;Residues: 1-103 <RYA>
 A;Cross-references: GB:M0299; NID:9108083; PIDN:AAA73873.1; PID:918084
 R;Su, M.W.; Lee, B.; Ramirez, F.; Machado, M.; Horton, W.
 Nucleic Acids Res. 17, 9473, 1989
 A;Title: Nucleotide sequence of the full length cDNA encoding for human type II proco
 A;Reference number: S06715; MUID:9006746
 A;Accession: S06715
 A;Molecule type: mRNA
 A;Residues: 1-28 <VRK>
 A;Cross-references: EMBL:X16468; NID:929515; PIDN:CAA34488.1; PID:929516
 A;Title: Alternative splice form 1
 R;Vikkula, M.; Metsaeranta, M.; Syraenen, A.C.; Ala-Kokko, L.; Vuorio, E.; Peltonen,
 Biochem. J. 285, 287-294, 1992
 A;Title: Structural analysis of the regulatory elements of the type-II procollagen ge
 A;Reference number: S24270; MUID:9334585
 A;Accession: S24270
 A;Status: translation not shown
 A;Molecule type: DNA
 A;Residues: 1-28 <VRK>
 A;Cross-references: EMBL:X58709; GB:S40537; NID:935659
 A;Note: this translation is not annotated in GenBank entry HSPROCOL1, release 11.0
 R;Nunez, A.M.; Kohno, K.; Martin, G.R.; Yamada, Y.
 Gene 44, 11-16, 1985
 A;Title: Promoter region of the human pro-alpha-1(II)-collagen gene.
 A;Reference number: A24828; MUID:87031574
 A;Accession: A24828
 A;Molecule type: DNA
 A;Residues: 1-8; T, 10-28 <NUUN>
 A;Cross-references: GB:M25698; NID:9108072; PIDN:AAA52051.1; PID:9553237
 R;Baldwin, C.T.; Regnato, A.M.; Smith, C.; Jimenez, S.A.; Prockop, D.J.
 Biochem. J. 262, 525-528, 1989
 A;Title: Structure of cDNA clones coding for human type II procollagen
 A;Reference number: S06496; MUID:90026318
 A;Accession: S06496
 A;Molecule type: mRNA
 A;Residues: 7-28, 'R', 99-157, 'P', 159-440, 'G', 442-456, 'E', 458-640, 'A', 642-831, 'PA', 834,
 A;Cross-references: EMBL:X16711; NID:930040; PIDN:CAA34683.1; PID:930041

A; Note: alternative splice form 1
 R; Ryan, M.C.; Sandell, L.J.
 J. Biol. Chem. 265, 10334-10339, 1990
 A; Title: Differential expression of a cysteine-rich domain in the amino-terminal propeptidase
 A; Reference number: A35428; MUID:90285153
 A; Accession: A35428
 A; Status: not compared with conceptual translation
 A; Molecule type: mRNA
 A; Residues: 27-81, 'L', 83-103 <RY2>
 A; Note: alternative splice form 2; splicing appears to be under developmental regulation
 R; Su, M.W.; Benson-Chanda, V.; Vissing, H.; Ramirez, F.
 Genomics 4, 438-441, 1989
 A; Title: Organization of the exons coding for Pro alpha-1(II) collagen N-propeptide confirmed by sequencing of cDNA clones
 A; Reference number: A30147; MUID:89233138
 A; Molecule type: DNA
 A; Residues: 104-157, 'P', 159-236 <SUM>
 A; Cross-references: GB:J03055; GB:M22660; GB:M25655; GB:M25656; GB:M25730; GB:M32168; GB:R; Ala-Kokko, L.; Baldwin, C.T.; Moskowitz, R.W.; Prockop, D.J.
 Proc. Natl. Acad. Sci. U.S.A. 87, 6563-6566, 1990
 A; Title: Single base mutation in the type II procollagen gene (COL2A1) as a cause of primary osteoarthritis
 A; Reference number: A94227; MUID:90370826
 A; Accession: A33116
 A; Molecule type: DNA
 A; Residues: 171-172, 'C', 174-175 <ALA>
 A; Note: mutant sequence from a family with family with primary generalized osteoarthritis
 R; Pfab, M.; Wu, J.J.; Eyre, D.R.
 Biochem. J. 314, 327-332, 1996
 A; Title: Collagen type IX from human cartilage: a structural profile of intermolecular cross-links
 A; Reference number: S6673; MUID:96195147
 A; Accession: S6674
 A; Molecule type: protein
 A; Residues: 188-189, 'X', 191-195; 1224-1230, 'X', 1232-1236 <DXA>
 R; Franc, S.; Marzin, E.; Boutillon, M.M.; Lafont, R.; Lechene de la Porte, P.; Herbage, Eur. J. Biochem. 234, 125-131, 1995
 A; Title: Immunohistochemical and biochemical analyses of 20000-25000-year-old fossil cartilage
 A; Reference number: S66514; MUID:96096730
 A; Accession: S66514
 A; Molecule type: protein
 A; Residues: 243-261; 575-590, 756-763, 'X', 765-779 <FRAS>
 R; Tiller, G.E.; Weis, M.A.; Poluncho, P.A.; Gruber, H.E.; Rimoin, D.L.; Cohn, D.H.; Eyre, Am. J. Hum. Genet. 56, 388-395, 1995
 A; Title: An RNA-splicing mutation (G>IVS20) in the type II collagen gene (COL2A1) in a
 A; Reference number: 138867; MUID:95150028
 A; Accession: 138867
 A; Status: preliminary; translated from GB/EMBL/DBJ
 A; Molecule type: DNA
 A; Residues: 440, 'G', 442-456, 'E', 458-460, 'P', 482-509 <TTLL>
 A; Cross-references: EMBL:U15195; NID:9557033; PIDN:AMB60370.1; PID:9557054
 R; Ramirez, F.
 A; Reference number: 138867
 A; Accession: 138867
 A; Molecule type: mRNA
 A; Residues: 501-676, 'A', 678-783, 'A', 785-831, 'PA', 834, 'F', 836-1214 <GRAM>
 A; Cross-references: EMBL:X13783; NID:930037; PIDN:CAA32030.1; PID:930050
 R; Viikula, M.; Peitonen, L.
 FEBS Lett. 250, 171-174, 1989
 A; Title: Structural analyses of the polymorphic area in type II collagen gene.
 A; Reference number: S05000; MUID:89323561
 A; Accession: S05000
 A; Molecule type: DNA
 A; Residues: 530-640, 'A', 642-785 <VIK2>
 A; Cross-references: EMBL:X10158; NID:929951; PIDN:CAA34278.1; PID:91335018; PIDN:CAA3427
 PIDN:CAA34283.1; PID:91335023; PIDN:CAA34284.1; PID:91335024
 R; Bogaert, R.; Tiller, G.E.; Weis, M.A.; Gruber, H.E.; Rimoin, D.L.; Cohn, D.H.; Eyre, I. J. Biol. Chem. 267, 22522-22526, 1992
 A; Title: An amino acid substitution (Gly853-->Glu) in the collagen alpha 1(II) chain prevents the propeptidase from cleaving the propeptides at the C-terminus
 A; Reference number: A44309; MUID:93054548
 A; Status: nucleic acid sequence not shown; not compared with conceptual translation
 A; Molecule type: DNA; mRNA
 A; Residues: 752-831, 'PA', 834, 'F', 836-1005, 'K', 1007-1036, 'Q', 1038-1052, 'E', 1054-1068, 'T', 1069-1109, 120
 A; Cross-references: GB:L00977; NID:9180812; PIDN:AB23914.1; PID:9258774
 A; Note: sequence extracted from NCBI backbone (NCBIP11723); parts of this sequence
 A; Note: this translation is not annotated and this publication is not cited in GenBank
 A; Note: mutant sequence associated with perinatal lethal hypochondroplasia
 R; Miller, G.E.; Rimoin, D.L.; Murray, L.W.; Cohn, D.H.
 Proc. Natl. Acad. Sci. U.S.A. 87, 3899-3893, 1990
 A; Title: Tandem duplication within a type II collagen gene (COL2A1) exon in an individual
 A; Reference number: S16502; MUID:90251662
 A; Molecule type: DNA
 A; Residues: 1164-1184, 'GPSKGDKANGSIPGPT', 1185-1199 <NTL2>
 A; Cross-references: EMBL:M37126; NID:9180808; PIDN:AA52037.1; PID:9180809
 A; Note: mutant sequence from a patient with spondyloepiphyseal dysplasia
 R; Cheah, K.S.B.; Stoker, N.G.; Griffin, J.R.; Grosveld, F.G.; Solomon, E.
 Proc. Natl. Acad. Sci. U.S.A. 82, 2555-2559, 1985
 A; Title: Identification and characterization of the human type II collagen gene (COL2A1)
 A; Reference number: A02858; MUID:85190534
 A; Accession: A02858
 A; Molecule type: DNA
 A; Residues: 1032-1056, 'N', 1058-1068, 'T', 1070-1487 <CHE>
 A; Cross-references: GB:J00116; NID:9180395; PIDN:AA51997.1; PID:9180396
 R; Elima, K.; Vuorio, T.; Vuorio, E.
 Nucleic Acids Res. 15, 9499-9504, 1987
 A; Title: Determination of the single polyadenylation site of the human pro-alpha-1(III) collagen gene
 A; Reference number: A27280; MUID:88067771
 A; Accession: A27280
 A; Molecule type: DNA; mRNA
 A; Cross-references: EMBL:X06268; NID:930096; PIDN:CAA29604.1; PID:930097
 A; Residues: 1175-1487 <ELI>
 A; Cross-references: EMBL:X06268; NID:930096; PIDN:CAA29604.1; PID:930097
 A; Experimental source: fetal epiphyseal cartilage
 R; van der Rest, M.; Rosenberg, L.C.; Olsen, B.R.; Poole, A.R.
 Biochem. J. 237, 923-925, 1986
 A; Title: Chondrocalcin is identical with the C-propeptide of type II procollagen.
 A; Reference number: A57033; MUID:87099927
 A; Accession: A57033
 A; Molecule type: protein
 A; Residues: 'XE', 1244-1245, 'N', 1248, 'X', 1250-1265; 1295-1305; 1395-1408 <VAN>
 A; Note: chondrocalcin identified as released collagen 1(II) chain carboxyl-terminal peptide
 R; Strom, C.M.; Upholt, W.B.
 Nucleic Acids Res. 12, 1025-1038, 1984
 A; Title: Isolation and characterization of genomic clones corresponding to the human
 A; Reference number: A21733; MUID:84118798
 A; Accession: A21733
 A; Molecule type: DNA
 A; Residues: 1245-1295 <STR1>
 A; Cross-references: EMBL:X00339; EMBL:X00298; NID:9394699; PIDN:CAA25092.1; PID:94378
 A; Accession: B21733
 A; Molecule type: DNA
 A; Residues: 894-909, 'PF', <STR2>
 A; Cross-references: EMBL:K0185; NID:930035; PIDN:CAA25082.1; PID:91335032
 R; Nunes, A.M.; Francomme, C.; Young, M.F.; Martin, G.R.; Yamada, Y.
 Biochemistry 24, 6343-6348, 1985
 A; Title: Isolation and partial characterization of genomic clones coding for a human gene.
 A; Reference number: A24561; MUID:86104139
 A; Accession: A24561
 A; Molecule type: DNA
 A; Residues: 1296-1358 <NUUN2>
 A; Cross-references: GB:W12048; NID:9180017
 A; Note: this translation is not annotated in GenBank entry HUMCCT2A, release 11.0
 A; Note: the codons given for 133-Lys (AGG) and 1350-Gly (GCA) are inconsistent with
 R; Sangiorgi, F.O.; Benson-Chanda, V.; de Wet, W.J.; Sobel, M.E.; Tsipouras, P.; Ramir
 Nucleic Acids Res. 13, 2207-2225, 1985
 A; Title: Isolation and partial characterization of the entire human pro alpha 1(II) c
 A; Reference number: I37249; MUID:85215609
 A; Accession: S55491
 A; Molecule type: DNA
 A; Residues: 7-28 <SAN2>
 A; Cross-references: GB:M23759; NID:9180845; EMBL:X03320; GB:M24938; NID:930104

A;Note: the GenBank PID is based on an incorrect reading frame
A;Accession: I17250
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 541-560 <SAN3>
A;Cross-references: EMBL:X02378; GB:M23870; NID:930107; PIDN:CAA26227.1; PID:929621
A;Accession: I17251

Query Match	59.8%	Score 55;	DB 1;	Length 1487;
Best Local Similarity	66.7%	Pred. No. 6;		
Matches	10;	Conservative	1;	Mismatches 4;
			Indels 0;	Gaps 0;

Qy 1 PAGPKWPGNGDKGKVG 15
Db 1165 PPGPVGPSPKGDKGANG 1179

RESULT 28

CGH2E
collagen alpha 2(XI) chain precursor - human (fragment)
N;Alternate names: procollagen alpha 2(XI) chain
N;Contains: proline/arginine-rich protein (PAPP)
C;Species: Homo sapiens (man)
C;Date: 07-Jun-1990 #sequence_revision 03-Oct-1995 #text_change 22-Jun-1999
C;Accession: S314790; A32645
R;Zhdikova, N.I.; Brewton, R.G.; Mayne, R.
FEBS Lett. 326, 25-28, 1993
A;Title: Molecular cloning of PAPP (proline/arginine-rich protein) from human cartilage
A;Reference number: S34790; MUID:9314796
A;Accession: S34790
A;Molecule type: mRNA
A;Residues: 1-663 <XH>
A;Cross-references: EMBL:118987; NID:9306439; PIDN:AAA35498.1; PID:9306440
R;Kimura, T.; Cheah, K.S.E.; Chan, S.D.H.; Lui, V.C.H.; Mattei, M.G.; van der Rest, M.;
J. Biol. Chem. 264, 13910-13916, 1989
A;Title: The human alpha2(XI) collagen (COL11A2) chain. Molecular cloning of cDNA and ge
A;Reference number: A32645; MUID:9340485
A;Molecule type: DNA; mRNA
A;Accession: A32645
A;Residues: 586-1546 <KIM>
A;Cross-references: GB:J04974; NID:9180714; PIDN:AAA52034.1; PID:9180715
A;Note: parts of this sequence were determined by protein sequencing
C;Comment: Prolines and lysines at the third position of the tripeptide repeating unit
ed and subsequently O-glycosylated.
C;Genetics:
A;Gene: GDB:COL11A2
A;Cross-references: GDB:119788; OMIM:120290
A;Map position: 6p21.3-6p21.3
A;Introns: 1302/3; 1320/3; 1350/3; 140/1; 1477/3
A;Note: the list of introns is incomplete
C;Complex: type XI collagen may be a heterotrimer of two alpha 1(XI) chains (see PIR:CGH
3(XI)) chain (see PIR:CGH06C), initially linked by disulfide bonds among their carboxyl-
rmed with desmosine cross-links made from lysine and allysine residues
C;Function:
A;Description: structural component of extracellular fibrous polymer associated with cell
A;Note: may play a role in controlling the lateral growth of collagen II fibrils
C;Superfamily: collagen alpha 1(V) chain; fibrillar collagen carboxyl-terminal homology
C;Keywords: coiled coil; extracellular matrix; glycoprotein; hydroxylysine; hydroxyproline
F:1-254/Domain: non-collagenous (fragment) #status predicted <NC3>
F:1-187/Product: proline-rich PAPP protein (fragment) #status predicted <PARP>
F:225-305/Domain: collagenous, triple helix #status predicted <COL2>
F:305-342/Domain: non-collagenous #status predicted <NC2>
F:343-1356/Region: helical
F:42-431/Region: cell attachment (R-G-D) motif
F:447-449/Region: cell attachment (R-G-D) motif
F:1125-1259/Region: cell attachment (R-G-D) motif
F:1357-1380/Region: carboxyl-terminal nonhelical telopeptide
F:1381-1546/Domain: carboxyl terminal propeptide (fragment) #status predicted <CTP>
F:140403-1546/Domain: fibrillar collagen carboxyl-terminal homology (fragment) #status aty
F:109-163 151-1545/Disulfide bonds: #status predicted
F:319/Modified site: allysine (Lys) #status predicted
F:426,1266/Modified site: 5-hydroxylysine (Lys) #status predicted

F:426,1266/Binding site: carbohydrate (Lys) (covalent) #status predicted
F:927, 933,1008,1017,1035,1038,1290,1296,1305,1317,1320/Modified site: 4-hydroxyprolin
F:929/Modified site: 4-hydroxyproline (Pro) #status atypical
F:942,1023,1299/Modified site: 5-hydroxylysine (Lys) #status experimental
F:942,11023,1299/Binding site: carbohydrate (Lys) (covalent) #status experimental
F:1427,1433,1450,1459/Disulfide bonds: interchain #status predicted
F:1460/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match	59.8%	Score 55;	DB 1;	Length 1546;
Best Local Similarity	66.7%	Pred. No. 6.3;		
Matches	10;	Conservative	0;	Mismatches 5;
			Indels 0;	Gaps 0;

Qy 1 PAGPKWPGNGDKGKVG 15
Db 1040 PRGPAGPNGADGPQG 1054

RESULT 29

A44982
collagen UC011 - pig roundworm (fragment)
C;Species: Ascaris suum (pig roundworm)
C;Date: 14-May-1993 #sequence_revision 14-May-1993 #text_change 20-Sep-1999
C;Accession: A44982
R;Kingston, I.B.; Weinwright, S.M.; Cooper, D.
Mol. Biochem. Parasitol. 37, 137-146, 1989
A;Title: Comparison of collagen gene sequences in Ascaris suum and Caenorhabditis ele
A;Reference number: A44982; MUID:90136710
A;Accession: A44982
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-228 <XIN>
A;Cross-references: GB:J04657; NID:9159656; PIDN:AAA29371.1; PID:9159657
C;Superfamily: unassigned collagens

Query Match	58.7%	Score 54;	DB 2;	Length 228;
Best Local Similarity	66.7%	Pred. No. 1.3;		
Matches	10;	Conservative	1;	Mismatches 4;
			Indels 0;	Gaps 0;

Qy 1 PAGPKWPGNGDKGKVG 15
Db 41 PPGPPGPDPGDKGDPDG 55

RESULT 30

T20177
hypothetical protein C53B4.5 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000
C;Accession: T20177
R;Barks, M.
submitted to the EMBL Data Library, December 1995
A;Reference number: Z19233
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-289 <NTL>
A;Cross-references: EMBL:Z68215; PIDN:CAA92453.1; GSPDB:GN00022; CESP:C53B4.5
A;Experimental source: clone C53B4
A;Genes:
C;Gene: CESP:C53B4.5
A;Map position: 4
C;Superfamily: unassigned collagens

Query Match	58.7%	Score 54;	DB 2;	Length 289;
Best Local Similarity	66.7%	Pred. No. 1.6;		
Matches	10;	Conservative	0;	Mismatches 5;
			Indels 0;	Gaps 0;

Qy 1 PAGPKWPGNGDKGKVG 15
Db 178 PAGPPGPDPGDKGDPDG 192

Search completed: November 1, 2002, 12:54:19

Fri Nov 1 13:09:33 2002

Job time : 15.5 secs

us-09-529-691a-3.rpr

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 1, 2002, 12:16:05 ; Search time 7 Seconds

82.970 Million cell updates/sec

Title: US-09-529-691A-3

Perfect score: 92

Sequence: 1 PAGPMGPNKGDKVG 15

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	66	71.7	1364	1 CA21_BOVIN
2	64	69.6	1356	1 CA21_ONCWT
3	63	68.5	1366	1 CA21_CANFA
4	63	68.5	1365	1 CA21_HUMAN
5	62	67.4	526	1 CA21_RABIT
6	60	65.2	1372	1 CA21_MOUSE
7	60	65.2	1372	1 CA21_RAT
8	60	65.2	1 CA17_HUMAN	
9	57	62.0	1459	1 CA12_MOUSE
10	56	60.9	369	1 CA12_CHICK
11	56	60.9	636	1 CA13_RAT
12	56	60.9	1804	1 CA1B_MOUSE
13	56	60.9	1806	1 CA1B_HUMAN
14	55	59.8	1262	1 CA13_CHICK
15	55	59.8	1362	1 CA21_CHICK
16	55	59.8	1418	1 CA21_HUMAN
17	55	59.8	1453	1 CA11_CHICK
18	55	59.8	1464	1 CA13_MOUSE
19	55	59.8	1650	1 CA2B_MOUSE
20	55	59.8	1736	1 CA2B_HUMAN
21	54	58.7	305	1 YK61_CAEEL
22	54	58.7	1049	1 CA13_BOVIN
23	54	58.7	1355	1 CA21_RANCA
24	54	58.7	1466	1 CA13_HUMAN
25	54	58.7	1527	1 CA1H_MOUSE
26	54	58.7	1838	1 CA15_HUMAN
27	53	57.6	307	1 CC36_CAEEL
28	53	57.6	329	1 CC05_CAEEL
29	53	57.6	374	1 PSPD_MOUSE
30	53	57.6	374	1 PSPD_RAT
31	53	57.6	779	1 CAL1_BOVIN
32	53	57.6	1496	1 CA25_HUMAN
33	56.5	301	1 CC02_CAEEL	

ALIGMENTS

RESULT 1	CA21_BOVIN	ID CA21_BOVIN STANDARD: PRT; 1364 AA.
CA21_BOVIN	ID CA21_BOVIN STANDARD: PRT; 1364 AA.	ID CA21_BOVIN STANDARD: PRT; 1364 AA.
09384	oncorhynchu	09384
046392	canis	046392
P08123	homo	P08123
P08668	oryctolagus	P08668
Q01149	mus	Q01149
P02466	rattus	P02466
P02388	homo	P02388
P28481	mus	P28481
P02460	gallus	P02460
P13341	rattus	P13341
Q61245	mus	Q61245
P12107	homo	P12107
P1205	gallus	P1205
P02467	gallus	P02467
P02458	homo	P02458
P02457	gallus	P02457
P08121	mus	P08121
P64739	mus	P64739
P11942	homo	P11942
P02467	gallus	P02467
RN	SEQUENCE OF 80-98.	SEQUENCE OF 80-98.
RC	TISSUE=SKIN;	TISSUE=SKIN;
RX	MEDLINE=75036115; PubMed=4609475;	MEDLINE=75036115; PubMed=4609475;
RA	Fierzek P.P., Breitkreitz D., Kuehn K.;	Fierzek P.P., Breitkreitz D., Kuehn K.;
RT	"Amino acid sequence of the amino-terminal region of calf skin collagen.";	"Amino acid sequence of the amino-terminal region of calf skin collagen.";
RT	Fierzek P.P., Rexrodt F.W.;	Fierzek P.P., Rexrodt F.W.;
RL	Biochim. Biophys. Acta 365:305-310(1974).	Biochim. Biophys. Acta 365:305-310(1974).
RN	[3]	[3]
RP	SEQUENCE OF 95-415, AND REVISION.	SEQUENCE OF 95-415, AND REVISION.
RC	TISSUE=SKIN;	TISSUE=SKIN;
RX	MEDLINE=76091874; PubMed=173531;	MEDLINE=76091874; PubMed=173531;
RA	Fierzek P.P., Rexrodt F.W.;	Fierzek P.P., Rexrodt F.W.;
RT	"The covalent structure of collagen. The amino-acid sequence of alpha2-CB4 from calf-skin collagen.";	"The covalent structure of collagen. The amino-acid sequence of alpha2-CB4 from calf-skin collagen.";
RL	Eur. J. Biochem. 59:113-118(1975).	Eur. J. Biochem. 59:113-118(1975).
RN	[4]	[4]
RC	SEQUENCE OF 416-445.	SEQUENCE OF 416-445.
RC	TISSUE=SKIN;	TISSUE=SKIN;
RX	MEDLINE=75008198; PubMed=4412529;	MEDLINE=75008198; PubMed=4412529;
RA	Fierzek P.P., Furthmayr H., Kuehn K.;	Fierzek P.P., Furthmayr H., Kuehn K.;
RT	"Comparative sequence studies on alpha2-CB2 from calf, human, rabbit and pig-skin collagen.";	"Comparative sequence studies on alpha2-CB2 from calf, human, rabbit and pig-skin collagen.";
RL	Eur. J. Biochem. 47:257-261(1974).	Eur. J. Biochem. 47:257-261(1974).
RN	[5]	[5]
RC	SEQUENCE OF 446-481.	SEQUENCE OF 446-481.
RC	TISSUE=SKIN;	TISSUE=SKIN;
RX	MEDLINE=75059350; PubMed=4435743;	MEDLINE=75059350; PubMed=4435743;
RA	Fierzek P.P., Kuehn K.;	Fierzek P.P., Kuehn K.;
RT	"The covalent structure of collagen: amino acid sequence of the N-terminal region of alpha2-CB3 from rat skin collagen and alpha2-CB3.5 from calf skin collagen.";	"The covalent structure of collagen: amino acid sequence of the N-terminal region of alpha2-CB3 from rat skin collagen and alpha2-CB3.5 from calf skin collagen.";
RL	Hoppe-Seyler's Z. Physiol. Chem. 355:647-650(1974).	Hoppe-Seyler's Z. Physiol. Chem. 355:647-650(1974).

CC -!- FUNCTION: TYPE I COLLAGEN IS A MEMBER OF GROUP I COLLAGEN
 CC (FIBRILLAR FORMING COLLAGEN).
 CC -!- SUBUNITS: TRIMERS OF ONE ALPHA 2(I) AND TWO ALPHA 1(I) CHAINS.
 CC -!- TISSUE SPECIFICITY: FORMS THE FIBRILS OF TENDON, LIGAMENTS AND
 CC BONES. IN BONES THE FIBRILS ARE MINERALIZED WITH CALCIUM
 CC HYDROXYLAPATE.
 CC -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
 UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@lsb-sib.ch).
 CC
 CC EMBL; AF035120; AC:64485; 1;
 DR InterPro; IPR00087; Collagen.
 DR InterPro; IPR00085; Fib_collagen_C.
 DR Pfam; PF01410; COL1; 1.
 DR Pfam; PF01391; Collagen; 18.
 DR Prodom; PD002078; Fib_collagen_C; 1.
 DR SMART; SM0038; COFI; 1.
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 KW Collagen; Signal; 1.
 FT SIGNAL 1 24 POTENTIAL,
 FT PROPEP 25 79 AMINO-TERMINAL PROPEPTIDE (BY
 FT CHAIN 80 1102 COLLAGEN SIMILARITY).
 FT PROPEP 1103 1366 CARBOXYL-TERMINAL PROPEPTIDE (BY
 FT MOD_RES 80 80 PYRROLIDONE CARBOXYLIC ACID (BY
 FT MOD_RES 84 84 SIMILARITY).
 FT INVOLVED TO AN ALDEHYDE GROUP THAT IS
 FT SIMILARITY).
 SQ SEQUENCE 1366 AA; 129400 MW; CD936969E080BCD4 CRC64;
 Query Match 68.5%; Score 63; DB 1; Length 1366;
 Best Local Similarity 73.3%; Pred. No. 0.33; [7]
 Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 Db 1055 PAGPSGPACKDGRTG 1069
 Oy 1 PAGPGPGNGKDGKV 15
 Db 1055 PAGPSGPACKDGRTG 1069
 RESULT 4
 CAR1_HUMAN
 ID CAR1_HUMAN STANDARD; PRT: 1366 AA.
 AC P08123; P02464; Q9UEB6; Q9UPH0;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Collagen alpha 2(I) chain precursor.
 GN COL1A2.
 OS Homo sapiens (Human).
 OC Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homindae; Homo.
 OX NCBI_TAXID=9606;
 RN [1] SEQUENCE FROM N.A.
 RP MEDLINE=88058462; PubMed=2824475;
 RN de Wet W.J., Benson M.P., Benson-Chanda V., Chu M.-L., Dickson L.A.,
 RA Weil D., Ramirez F.;
 RT "Organization of the human pro-alpha 2(I) collagen gene.",
 RL J. Biol. Chem. 262:16032-16036(1987).
 RN [2] SEQUENCE FROM N.A.
 RP Korkko J.M., Earley J.J., Ala-Kokko L., Prockop D.J.;
 RA "Analysis of the COL1 and COL1A2 genes by CSGE and DNA sequencing in
 RT 14 patients with mild OI (Type I). Identification of common sequences
 RT for null allele mutations.",
 RT Submitted (MAY-1997) to the EMBL/GenBank/DDJB databases.
 RN [3] SEQUENCE OF 1-765 FROM N.A.
 RC TISSUE=Placenta;
 RC MEDLINE=88339824; PubMed=3421913;
 RA Kuijaniemi H., Tromp G., Chu M.-L., Prockop D.J.;
 RT "Structure of a full-length cDNA clone for the prepro alpha 2(I)
 chain of human type I procollagen. Comparison with the chicken gene
 RT confirms unusual patterns of gene conservation.";
 RL Biochem. J. 252:633-640(1988).
 RN [4] SEQUENCE OF 181-1366 FROM N.A.
 RA Kalicki J., Wamsley P., Gibson A.;
 RT Submitted (SEP-1997) to the EMBL/GenBank/DDJB databases.
 RN [5] SEQUENCE OF 623-1366 FROM N.A.
 RA MEDLINE=83178319; PubMed=6687691;
 RA Bernard M.P., Myers J.C., Chu M.-L., Ramirez F., Eikenberry E.F.,
 RA Prockop D.J.;
 RT "Structure of a cDNA for the pro alpha 2 chain of human type I
 procollagen. Comparison with chick cDNA for pro alpha 2(I) identifies
 RT structurally conserved features of the protein and the gene.";
 RL Biochemistry 22:1139-1145(1983).
 RN [6] SEQUENCE OF 80-96.
 RP TISSUE=Skin;
 RX MEDLINE=71038625; PubMed=5529814;
 RA Click E.M., Bernstein P.;
 RT "Isolation and characterization of the cyanoen bromide peptides from
 RT the alpha 1 and alpha 2 chains of human skin collagen.";
 RL Biochemistry 9:4699-4706(1970).
 RN [7] SEQUENCE OF 417-447.
 RP TISSUE=Skin;
 RC MEDLINE=75008198; PubMed=4412529;
 RA Fietzek P.P., Furthmayr H., Kuehn K.;
 RT "Comparative sequence studies on alpha2-CB2 from calf, human, rabbit
 RT and pig-skin collagen.";
 RL Eur. J. Biochem. 47:257-261(1974).
 RN [8] SEQUENCE OF 145-198 FROM N.A.
 RP MEDLINE=88298792; PubMed=3403536;
 RA Kuijaniemi H., Sabol C., Tromp G., Sippola Thiele M., Prockop D.J.;
 RT "A 19-base pair deletion in the pro-alpha 2(I) gene of type I
 RT procollagen that causes in-frame RNA splicing from exon 10 to exon 12
 RT in a proband with atypical osteogenesis imperfecta and in his
 RT asymptomatic mother.";
 RL J. Biol. Chem. 263:11407-11413(1988).
 RN [9] SEQUENCE OF 960-1351 FROM N.A.
 RC TISSUE=Skin;
 RX MEDLINE=90304220; PubMed=2364107;
 RA Maeelae J.K., Vuorio T., Vuorio E.;
 RT "Growth-dependent modulation of type I collagen production and mRNA
 RT levels in cultured human skin fibroblasts.";
 RL Biochim. Biophys. Acta 1049:171-176(1990).
 RN [10] REVIEW ON VARIANTS.
 RP MEDLINE=91184577; PubMed=2010058;
 RA Kuijaniemi H., Tromp G., Prockop D.J.;
 RT "Mutations in collagen genes: causes of rare and some common diseases
 RT in humans";
 RL FASEB J. 5:2052-2060(1991).
 RN [11] REVIEW ON VARIANTS.
 RP MEDLINE=9725559; PubMed=9101290;
 RA Kuijaniemi H., Tromp G., Prockop D.J.;
 RT "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-
 associated collagen (type IX), and network-forming collagen (type X)
 RT cause a spectrum of diseases of bone, cartilage, and blood vessels.";
 RL Hum. Mutat. 9:300-315(1997).
 RN [12]

RP REVIEW ON OI VARIANTS.

RX MEDLINE=91374476; PubMed=1695312;

RA Byers P.H., Wallis G.A., Willing M.C.;

RT "Osteogenesis imperfecta: translation of mutation to phenotype.";

RL J. Med. Genet. 28:433-442(1991).

RN [13]

RP REVIEW ON OI VARIANTS.

RX MEDLINE=9159389; PubMed=9016532;

RA Dalglish R.;

RT "The human type I collagen mutation database.";

RL Nucleic Acids Res. 25:181-187(1997).

RN [14]

RP VARIANT EDS-VII-A2.

RX MEDLINE=88059013; PubMed=3600255;

RA Wirtz M.K., Gianville R.W., Steinmann B., Rao V.H., Hollister D.W.;

RT "Ehlers-Danlos syndrome type VII B. Deletion of 18 amino acids comprising the N-terminal peptide region of a pro-alpha 2(I) chain.";

RL J. Biol. Chem. 262:16376-16385(1987).

RN [15]

RP SEQUENCE OF 1900-1107 FROM N.A., AND VARIANT OI-IV ARG-1102.

RX MEDLINE=88227975; PubMed=2897363;

RA Wenstrup R.J., Cohn D.H., Cohen T., Byers P.H.;

RT "Arginine for glycine substitution in the triple-helical domain of the products of one alpha 2(I) collagen allele (COLIA2) produces the osteogenesis imperfecta type IV phenotype.";

RL J. Biol. Chem. 263:7734-7740(1988).

RN [16]

RP VARIANT OI-II ASP-997.

RX MEDLINE=89123407; PubMed=2914942;

RA Baldwin C.T., Constantino C., Dumars K.W., Prockop D.J.;

RT "A single base mutation that converts glycine 907 of the alpha 2(I) chain of type I procollagen to aspartate in a lethal variant of osteogenesis imperfecta. The single amino acid substitution near the carboxyl terminus destabilizes the whole triple helix.";

RL J. Biol. Chem. 264:3002-3006(1989).

RN [17]

RP VARIANT OI-II SER-955.

RX MEDLINE=89380165; PubMed=277764;

RA Lamande S.R., Dahl H.-H.M., Cole W.G., Bateman J.F.;

RT "Characterization of point mutations in the collagen COLIA1 and COLIA2 genes causing lethal perinatal osteogenesis imperfecta.";

RL J. Biol. Chem. 264:15809-15812(1989).

RN [18]

RP VARIANT OI-II CIS-877.

RX Fertala A., Westerhausen A., Morris G.M., Rooney J.E., Prockop D.J.; "Two cysteine substitutions in the type I procollagen genes (COLIA1 and COLIA2) that cause lethal osteogenesis imperfecta. The location of glycine substitutions does not in any simple way predict their effects on protein function or phenotype.";

RL Ann. J. Hum. Genet. 47:A216-A216(1990).

RN [19]

RP VARIANT EDS-VII-A2.

RX MEDLINE=9038825; PubMed=2394758;

RA Weill D., Alessio M., Ramirez F., Eyre D.R.;

RT "Structural and functional characterization of a splicing mutation in the pro-alpha 2(I) collagen gene or an Ehlers-Danlos type VII patient.";

RL J. Biol. Chem. 265:16007-16011(1990).

RN [20]

RP VARIANT OI-IV VAL-676.

RX MEDLINE=91291136; PubMed=2064612;

RA Bateman J.F., Hannigan M., Chan D., Cole W.G.;

RT "Characterization of a type I collagen alpha 2(I) glycine-586 to valine substitution in osteogenesis imperfecta type IV. Detection of the mutation and prenatal diagnosis by a chemical cleavage method.";

RL Biochem. J. 276:765-770(1991).

RN [21]

RP VARIANTS OF CYS-349 AND CYS-736.

RX MEDLINE=911589; PubMed=199009;

RA Wenstrup R.J., Shirago-Howe A.W., Lever L.W., Phillips C.L., Byers P.H., Cohn D.H.;

RT "The effects of different cysteine for glycine substitutions within alpha 2(I) chains. Evidence of distinct structural domains within the

RP type I collagen triple helix.";

RL J. Biol. Chem. 266:2590-2594(1991).

RN [22]

RP VARIANT OI-II ARG-784.

RX MEDLINE=91340689; PubMed=1874719;

RA Spotila L.D., Constantinou C.D., Sereda L., Ganguly A., Riggs B.L., Prockop D.J.;

RT "Mutation in a gene for type I procollagen (COLIA2) in a woman with postmenopausal osteoporosis: evidence for phenotypic and genotypic overlap with mild osteogenesis imperfecta.";

RL Proc. Natl. Acad. Sci. U.S.A. 88:5423-5427(1991).

RN [23]

RP VARIANT OI-IV SER-751.

RX MEDLINE=91271401; PubMed=2052622;

RA Spotila L.D., Constantinou C.D., Sereda L., Ganguly A., Riggs B.L., Prockop D.J.;

RT "Arginine for glycine substitution detected by chemical cleavage of alpha 2(I) gly to Arg substitution mismatch.";

RL Human Mutat. 1:55-62(1992).

RN [24]

RP VARIANT OI-II ARG-547.

RX MEDLINE=93244832; PubMed=1284475;

RA Bateman J.F., Moeller T., Hannigan M., Chan D., Cole W.G.;

RT "Lethal perinatal osteogenesis imperfecta due to a type I collagen alpha 2(I) Gly to Arg substitution mismatch.";

RL Human Mutat. 1:55-62(1992).

RN [25]

RP VARIANT OI-II ASP-670.

RX MEDLINE=93054637; PubMed=1385413;

Db 1055 PAGSPGPKDQRTG 1069

OY RESULT 5

CA21_RABBIT

ID	CA21_RABBIT	STANDARD	PRF:	526 AA.
AC	028668;			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Collagen alpha 2(I) chain precursor (Fragment).			
GN	COLIA2.			
OS	Oryctolagus cuniculus (Rabbit).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.			
OX	NCBI_TAXID=986;			
RN	[1] SEQUENCE FROM N.A.			
RP	STRAIN="New Zealand white; TISSUE=Calvaria;			
RA	Inoue S., Okazaki T.;			
RT	"Alpha 2 type I collagen gene expression in the rabbit knee ligaments: variations during the newborn development and in the adult age.";			
RL	Biomol. Res. 16:219-227(1995).			
CC	-1- FUNCTION: TYPE I COLLAGEN IS A MEMBER OF GROUP I COLLAGEN (FIBRILLAR FORMING COLLAGEN).			
CC	-1- SUBUNIT: TRIMERS OF ONE ALPHA 2(I) AND TWO ALPHA 1(I) CHAINS.			
CC	-1- TISSUE SPECIFICITY: FORMS THE FIBRILS OF TENDON, LIGAMENTS AND BONES. IN BONES THE FIBRILS ARE MINERALIZED WITH CALCIUM HYDROXYAPATITE.			
CC	-1- PWW: PROLINES AT THE THIRD POSITION OF THE TRIPETIDE REPEATING UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.			

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@sb-sib.ch).

EMBL: DR|3999; BAAB8391; -
 DR|InterPro; IPR00007; Collagen.
 DR|InterPro; IPR000885; Fib_collagen_C.
 DR|Pfam; PF01410; COLFI; 1.
 DR|Prodom; PD002078; Fib_collagen_C; 1.
 DR|PROTEIN; SN00038; COLFI; 1.
 KW|Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 KW|Collagen.
 NON_TER|1 1 262
 CHAIN|<1 1 526
 PROPEP|263 526
 SEQUENCE| 526 AA; 53129 MW; 7CA8FORA6953846D CRC64;
 Query Match| 67.4%; Score 62; DB 1; Length 526;
 Best Local Similarity| 73.3%; Pred. No. 0.19;
 Matches| 11; Conservative| 1; Mismatches| 3; Indels| 0; Gaps| 0;
 QY| 1 PAGPWGPNKGDKGVG 15
 Db| 215 PAGPTGPAGKDGSG 229

RESULT 6

ID CA21_MOUSE	STANDARD:	PRT: 1372 AA.
AC 001149;	CREATED:	
DT 01-APR-1993 (Rel. 25, Created)	UPDATED:	
DT 16-OCT-2001 (Rel. 40, Last sequence update)	LAST ANNOTATION:	
DE Collagen alpha 2(I) chain precursor.		
GN COLIA2 OR COLA2.		
OS Mus musculus (Mouse).		
OC Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Buteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX NCBI_TAXID=10090;		
RN [1]		
RP SEQUENCE FROM N.A.		
RC TISSUE=Salvaria;		
RX MEDLINE=92372043; PubMed=1505972;		
RA Phillips C.L., Morgan A.L., Lever L.W., Wenstrup R.J.;		
RT "Sequence analysis of a full-length cDNA for the murine pro alpha 2(I) collagen chain: comparison of the derived primary structure with human pro alpha 2(I) collagen.";		
RL Genomics 13:1345-1346(1992). [2]		
RN [2]		
RP SEQUENCE FROM N.A.		
RC TISSUE=breast tumor;		
RA Strausberg R.;		
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.		
RN [3]		
RP SEQUENCE OF 1-110 FROM N.A.		
RC TISSUE=Salvaria;		
RX MEDLINE=92084969; PubMed=1748823;		
RA Philipps C.L., Lever L.W., Pinnell S.R., Quarles L.D., Wenstrup R.J.;		
RT "Construction of a full-length murine pro alpha 2(I) collagen cDNA by the polymerase chain reaction.";		
RL J. Invest. Dermatol. 97:980-984(1991). [4]		
RP SEQUENCE OF 1-23 FROM N.A.		
RX MEDLINE=97289650; PubMed=3039494;		
RA Rossi P.; de Crombrugge B.;"Identification of a cell-specific transcriptional enhancer in the first intron of the mouse alpha 2 (type I) collagen gene.";		
RT Proc. Natl. Acad. Sci. U.S.A. 84:5590-5594(1987). [5]		
RL -I- FUNCTION: TYPE I COLLAGEN IS A MEMBER OF GROUP I COLLAGEN		

CC| (FIBRILLAR FORMING COLLAGEN).
 CC| -- SUBUNIT: TRIMERS OF ONE ALPHA 2(I) AND TWO ALPHA 1(I) CHAINS.
 CC| -- TISSUE SPECIFICITY: FORMS THE FIBRILS OF TENDON, LIGAMENTS AND
 CC| BONES. IN BONES THE FIBRILS ARE MINERALIZED WITH CALCIUM
 CC| HYDROXYAPATITE.
 CC| -- PTM: PROLINES AT THE THIRD POSITION OF THE TRipeptide REPEATING
 CC| UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 CC| ---
 CC| This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC| between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC| the European Bioinformatics Institute. There are no restrictions on its
 CC| use by non-profit institutions as long as its content is in no way
 CC| modified and this statement is not removed. Usage by and for commercial
 CC| entities requires a license agreement (see <http://www.isb-sib.ch/announce/>,
 CC| or send an email to license@sb-sib.ch).
 CC| ---
 CC| EMBL: X88251; CAA41205; 1; -
 DR|EMBL; K01832; AAA37331; -.
 DR|PIR: A43291; A43291.
 DR|MGI: 88468; Col1a2.
 DR|InterPro; IPR000087; Collagen.
 DR|ProDom; PD002078; Fib_collagen_C; 1.
 DR|InterPro; IPR000885; Fib_collagen_C.
 DR|Pfam; PF01410; COLFI; 1.
 DR|Smart; SN00038; COLFI; 1.
 KW|Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 KW|Glycoprotein; Collagen; Signal.
 FT|SIGNAL| 1 22 POTENTIAL.
 FT|PROPEP| 23 85 AMINO-TERMINAL PROPEPTIDE
 FT| (BY SIMILARITY).
 FT|CHAIN| 86 1108 COLLAGEN ALPHA 2(I) CHAIN.
 FT|PROPEP| 1109 1372 CARBOXYL-TERMINAL PROPEPTIDE
 FT|MOD_RES| 86 86 PYRROLIDONE CARBOXYLIC ACID (BY
 FT|CONFLICT| 1273 1273 SIMILARITY).
 FT|MOD_RES| 90 90 CONVERTED TO AN ALDEHYDE GROUP THAT IS
 FT|CONFLICT| 1273 1273 INVOLVED IN CROSS-LINKING
 FT|SEQUENCE| 1372 AA; 129557 MW; 0D17DF5D6C1452D1 CRC64;
 Query Match| 65.2%; Score 60; DB 1; Length 1372;
 Best Local Similarity| 73.3%; Pred. No. 0.84;
 Matches| 11; Conservative| 1; Mismatches| 3; Indels| 0; Gaps| 0;
 QY| 1 PAGPWGPNKGDKGVG 15
 Db| 1061 PAGPSGPVKGDRSG 1075

RESULT 7

ID CA21_RAT	STANDARD:	PRT: 1372 AA.
AC P02466; Q9R1E8;	CREATED:	
DT 21-JUL-1986 (Rel. 01, Created)	UPDATED:	
DT 16-OCT-2001 (Rel. 40, Last sequence update)	LAST ANNOTATION:	
DE Collagen alpha 2(I) chain precursor.		
GN COLIA2.		
OS Rattus norvegicus (Rat). Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.		
OC NCBI_TAXID=10116;		
RN [1]		
RP SEQUENCE FROM N.A.		
RA Guenther D., Seibold S., Marx M.;"Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases." [2]		
RL SEQUENCE OF 86-98.		

RC TISSUE=Skin;
 RX MEDLINE=6712628; PubMed=5337885; Fib-collagen_C.
 RA Kang A.H.; Bornstein P.; Piez K.A.;
 RT "The amino acid sequence of peptides from the cross-linking region of
 rat skin collagen.";
 RL Biochemistry 67:88-95 (1967).
 [3]
 RP SEQUENCE OF 99-102.
 TISSUE=Skin;
 RX MEDLINE=65206881; PubMed=5785232;
 RA Fietzek P.P.; Piez K.A.;
 RT "Isolation and characterization of the cyanogen bromide peptides from
 the alpha 2 chain of rat skin collagen.";
 RL Biochemistry 8:2129-2133(1969).
 [4]
 RP SEQUENCE OF 102-144.
 TISSUE=Skin;
 RX MEDLINE=73049496; PubMed=4636752;
 RA Fietzek P.P.; Kell I.; Kuehn K.;
 RT "The covalent structure of collagen. Amino acid sequence of the N-
 terminal region of alpha 2-CB4 from calf and rat skin collagen.";
 RL FEBS Lett. 26:66-68(1972).
 [5]
 RP SEQUENCE OF 423-452.
 TISSUE=Skin;
 RX MEDLINE=7115126; PubMed=5541653;
 RA Highberger J.H.; Gross J.;
 RT "Comparative studies on the amino acid sequence of the alpha 2-CB2
 peptides from chick and rat skin collagens.";
 RL Biochemistry 10:610-616(1971).
 [6]
 RP SEQUENCE OF 453-501.
 TISSUE=Skin;
 RX MEDLINE=75059250; PubMed=4435743;
 RA Fietzek P.P.; Kuehn K.;
 RT "The covalent structure of collagen: amino acid sequence of the N-
 terminal region of alpha 2-CB3 from rat skin collagen and alpha 2-CB3.5
 from calf skin collagen";
 RT Hoppe-Seyler's Z. Physiol. Chem. 355:647-650(1974).
 RN [7]
 RP SEQUENCE OF 791-836.
 TISSUE=Skin;
 RX MEDLINE=74055004; PubMed=4763308;
 RA Fietzek P.P.; Kuehn K.;
 RT "The covalent structure of collagen: amino acid sequence of the N-
 terminal region of alpha 2-CB5 from rat skin collagen.";
 RL FEBS Lett. 36:289-291(1973).
 RN [8]
 RP ORDER OF CNBR PEPTIDES.
 RX MEDLINE=70181822; PubMed=5443712;
 RA Vuust J.; Lane J.M.; Fietzek P.P.; Miller E.J.; Piez K.A.;
 RT "The order of the CNBr peptides from the alpha 2 chain of collagen.";
 RL Biochem. Biophys. Res. Commun. 38:703-708(1970).
 CC -!- FUNCTION: TYPE I COLLAGEN IS A MEMBER OF GROUP I COLLAGEN
 CC (FIBRILLAR FORMING COLLAGEN).
 CC -!- SUBUNIT: TRIMERS OF ONE ALPHA 2(I) AND TWO ALPHA 1(I) CHAINS.
 CC -!- TISSUE SPECIFICITY: FORMS THE FIBRILS OF TENDON, LIGAMENTS AND
 CC BONES. IN BONES THE FIBRILS ARE MINERALIZED WITH CALCIUM
 CC HYDROXYAPATITE.
 CC -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPETIDE REPEATING
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 CC -----
 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 DR EMBL; AF021217; AAP01775.1;
 DR PIR; A02867; CGRT25.
 DR InterPro; IPR000087; Collagen.

DR IntePro; IPR00885; Fib-collagen_C.
 DR Pfam; PF01391; Collagen; 18.
 DR ProDom; PD002078; Fib.collagen_C; 1.
 DR SMART; SM0038; COLEI; 1.
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 KW Glycoprotein; Collagen; Signal.
 FT SIGNAL 1 24 POTENTIAL.
 FT PROPEP 25 AMINO-TERMINAL PROPEPTIDE
 FT CHAIN 85 (BY SIMILARITY; COLLAGEN ALPHA 2(I) CHAIN)
 FT PROPEP 1109 CARBOXY-TERMINAL PROPEPTIDE
 FT MOD_RES 86 CONVERTED TO AN ALDEHYDE GROUP THAT IS
 FT MOD_RES 90 INVOLVED IN CROSS-LINKING.
 FT CARBOYD 1273 N-LINED (GLCNAC...)(POTENTIAL).
 FT 132 T -> P (IN REF. 4).
 FT CONFLICT 132 S -> P (IN REF. 4).
 FT CONFLICT 137 MISSING (IN REF. 4).
 FT CONFLICT 145 ST -> TS (IN REF. 5).
 FT CONFLICT 432 E -> Z (IN REF. 6).
 FT CONFLICT 494 N -> A (IN REF. 6).
 FT CONFLICT 502 MISSING (IN REF. 6).
 FT CONFLICT 825 R -> K (IN REF. 7).
 FT CONFLICT 825 825 CRC64;
 SO SEQUENCE 1372 AA; 129564 MW; B069371A8DB20A72
 Query Match 65.2%; Score 60; DB 1; Length 1372;
 Best Local Similarity 73.3%; Pred. No. 0.84; 3; Indels 0; Gaps 0;
 Matches 11; Conservative 1; Mismatches 3;
 Qy 1 PAGPWNPNKGKV 15
 Db 1061 PAGPSPGPIGDGRSG 1075

RESULT 8
 CA17_HUMAN
 ID CA17_HUMAN
 AC 002388; Q14054; Q16507;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-FEB-1995 (Rel. 33, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Collagen alpha 1(VII) chain precursor (Long-chain collagen) (LC
 DE collagen).
 GN COL7A1.
 OS Homo sapiens (Human).
 OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarhini; Hominidae; Homo.
 OC Hominidae; Homo.
 OX NCBI_TAXID=9606;
 RP SEQUENCE FROM N. A.
 RX MEDLINE=94032750; PubMed=8051117;
 RA Christiano A.M.; Greenspan D.S.; Lee S.; Utton J.;
 RT "Cloning of human type VII collagen. Complete primary sequence of the
 alpha 1(VII) chain and identification of intragenic polymorphisms.";
 RL J. Biol. Chem. 269:20256-20262(1994).
 RN [2]
 RP SEQUENCE OF 120-1493 FROM N.A., AND PARTIAL SEQUENCE.
 RX MEDLINE=93138417; PubMed=1307247;
 RA Christiano A.M.; Rosenblum L.M.; Chung-Honet L.C.; Parente M.G.,
 RA Woodley D.T.; Pan T.C.; Zhang R.Z.; Chu M.-L.; Burgesson R.E.,
 RA Utton J.;
 RT "The large non-collagenous domain (NC-1) of type VII collagen is
 RT amino-terminal and chimeric. Homology to cartilage matrix protein,
 RT the type III domains of fibronectin and the A domains of von
 RT Willebrand factor.";
 RL Hum. Mol. Genet. 1:475-481(1992).
 RN [3]
 RP SEQUENCE OF 815-1439 FROM N.A.
 RX MEDLINE=91334380; PubMed=1871109;

RA Parente M.G., Chung L.C., Rynnaenen J., Woodley D.T., Wynn K.W.,
 RA Bauer E.A., Mattei M.-G., Chu M.-L., Utton J.;
 RT "Human type VII collagen: cDNA cloning and chromosomal mapping of the
 gene.," Proc. Natl. Acad. Sci. U.S.A. 88:6931-6935(1991).
 RN [4]

SEQUENCE OF 369-1255 FROM N.A.
 RX MEDLINE=93107742; PubMed=169284;
 RA Gaumon W.R., Abernethy M.L., Padilla K.M., Prisayanh P.S.,
 RA Cook M.E., Wright J., Brigham R.A., Hunt S.W. III;
 RT "Noncollagenous (NCI) domain of collagen VII resembles multidomain
 adhesion proteins involved in tissue-specific organization of
 extracellular matrix.," J. Invest. Dermatol. 99:691-696(1992).
 RL [5]

SEQUENCE OF 34-0-675 FROM N.A.
 RC TISSUE="keratinocytes";
 RX MEDLINE=92231902; PubMed=1567409;
 RA Tanaka T., Takahashi K., Furukawa F., Imamura S.;
 RT "Molecular cloning and characterization of type VII collagen cDNA.,"
 RL Biochem. Biophys. Res. Commun. 183:958-963(1992).
 RN [6]

SEQUENCE OF 2395-2944 FROM N.A.
 RX MEDLINE=93271985; PubMed=8199916;
 RA Greenspan D.S.,
 RT "The carboxyl-terminal half of type VII collagen, including the non-
 collagenous NC-2 domain and intron/exon organization of the
 corresponding region of the COL7A1 gene.,"
 RL Hum. Mol. Genet. 2:273-278(1993).
 RN [7]

SEQUENCE OF 1-87 FROM N.A.
 RC TISSUE="placenta";
 RX MEDLINE=94375010; PubMed=8088784;
 RA Christiano A.M., Hoffman G.G., Chung-Honet L.C., Lee S., Cheng W.,
 RA Utton J., Greenspan D.S.;
 RT "Structural organization of the human type VII collagen gene (COL7A1),
 composed of more exons than any previously characterized gene.,"
 RL Genomics 21:169-179(1994).
 RN [8]

REVIEW ON DEB VARIANTS.
 RX MEDLINE=98041696; PubMed=9375848;
 RA Jaakkilainen A., Puikkonen L., Utton J.;
 RT "Molecular basis of dystrophic epidermolysis bullosa: mutations in
 the type VII collagen gene (COL7A1).,"
 RL Hum. Mutat. 10:338-347(1997).
 RN [9]

VARIANT RDEB LYS-2798.
 RX MEDLINE=93291877; PubMed=8513326;
 RA Christiano A.M., Greenspan D.S., Hoffman G.G., Zhang X., Tamai Y.,
 RA Lin A.N., Dietz H.C., Hovnanian A., Utton J.;
 RT "A missense mutation in type VII collagen in two affected siblings
 with recessive dystrophic epidermolysis bullosa.,"
 RL Natl. Genet. 4:62-66(1993).
 RN [10]

VARIANT DDEB SER-2040.
 RX MEDLINE=94224777; PubMed=8170945;
 RA Christiano A.M., Rynnaenen M., Utton J.;
 RT "Dominant dystrophic epidermolysis bullosa: identification of a
 Gly->Ser substitution in the triple helical domain of type VII
 collagen.," Proc. Natl. Acad. Sci. U.S.A. 91:3549-3553(1994).
 RN [11]

VARIANT PEB-DDEB CYS-2623.
 RX MEDLINE=96081220; PubMed=8541842;
 RA Christiano A.M., Lee J.Y.-Y., Chen W.J., Laforgia S., Utton J.;
 RT "Pretilbial epidermolysis bullosa: genetic linkage to COL7A1 and
 identification of a glycine-to-cysteine substitution in the triple-
 helical domain of type VII collagen.," Hum. Mol. Genet. 4:1579-1583(1995).
 RN [12]

VARIANT DDEB ARG-2043.
 RX MEDLINE=95164985; PubMed=7861014;
 RA Christiano A.M., Morricone A., Paradisi M., Angelo C., Mazzanti C.,
 RN

Cavaliere R., Utton J.;
 RA "A glycine-toarginine substitution in the triple-helical domain of
 type VII collagen in a family with dominant dystrophic epidermolysis
 bullosa.," J. Invest. Dermatol. 104:438-440(1995).
 RN [13]

VARIANTS DDEB AND DDEB.
 RX MEDLINE=96220518; PubMed=8644729;
 RA Christiano A.M., McGrath J.A., Tan K.C., Utton J.;
 RT "Glycine substitutions in the triple-helical region of type VII
 collagen result in a spectrum of dystrophic epidermolysis bullosa
 phenotypes and patterns of inheritance.," Am. J. Hum. Genet. 58:671-681(1996).
 RL [14]

VARIANT RDEB ARG-2575.
 RX MEDLINE=9615068; PubMed=8592061;
 RA Shinilzu H., McGrath J.A., Christiano A.M., Nishikawa T., Utton J.;
 RT "Molecular basis of recessive dystrophic epidermolysis bullosa:
 genotype/phenotype correlation in a case of moderate clinical
 severity.," J. Invest. Dermatol. 106:119-124(1996).
 RL [15]

VARIANT RDEB ARG-1782.
 RX MEDLINE=96183562; PubMed=8618018;
 RA Christiano A.M., McGrath J.A., Utton J.;
 RT "Influence of the second COL7A1 mutation in determining the
 phenotype severity of recessive dystrophic epidermolysis bullosa.,"
 J. Invest. Dermatol. 106:766-770(1996).
 RN [16]

VARIANT RDEB ASP-2073.
 RX MEDLINE=96310789; PubMed=8757758;
 RA Dunnill M.G.S., McGrath J.A., Richards A.J., Christiano A.M.,
 RA Utton J., Pope F.M., Eddy R.A.J.;
 RT "Clinicopathological correlations of compound heterozygous COL7A1
 mutations in recessive dystrophic epidermolysis bullosa.,"
 RL J. Invest. Dermatol. 107:171-177(1996).
 RN [17]

VARIANTS RDEB W-1982; G-2008; A-2025; B-2049; G-2063 AND
 RX MEDLINE=97465605; PubMed=9326325;
 RA Hovnanian A., Rochat A., Bodemer C., Petit E., Rivers C.A., Prost C.,
 RA Fratigl S., Christiano A.M., Utton J., Lathrop M., Barrandon Y.,
 RA de Post Y.;
 RT "Characterization of 18 new mutations in COL7A1 in recessive
 RT dystrophic epidermolysis bullosa provides evidence for distinct
 RT molecular mechanisms underlying defective anchoring fibril
 RT formation.,"
 RL Am. J. Hum. Genet. 61:599-610(1997).
 RN [18]

VARIANT RDEB ARG-1652.
 RX MEDLINE=98106792; PubMed=9444387;
 RA Cserhalmi-Kriedman P.B., Karpati S., Horvath A., Christiano A.M.;
 RT "Identification of a glycine substitution and a splice site mutation
 RT in the type VII collagen gene in a proband with mitis recessive
 RT dystrophic epidermolysis bullosa.," Arch. Dermatol. Res. 289:640-645(1997).
 RN [19]

VARIANTS DEB ARG-2009 AND ARG-2043.
 RX MEDLINE=97350588; PubMed=9215684;
 RA Wimberg J.-O., Hammami-Haussi N., Nilissen O., Anton-Lamprecht I.,
 RA Navjor S.L., Kerbacher K., Zimmermann M., Krafcik P.,
 RA Gedde-Dahl T. Jr., Bruckner-Tuderman L.;
 RT "Modulation of disease severity of dystrophic epidermolysis bullosa by
 RT a splice site mutation in combination with a missense mutation in the
 RT COL7A1 gene.," Hum. Mol. Genet. 6:1125-1135(1997).
 RL [20]

VARIANTS DDEB ASP-1519; ASP-2006; GLU-2015 AND ARG-2034.
 RX MEDLINE=98334662; PubMed=968111;
 RA Hammami-Haussi N., Schumann H., Raghunath M., Kilgus O., Luethi U.,
 RA Luger T., Bruckner-Tuderman L.;
 RT "Some, but not all, glycine substitution mutations in COL7A1 result in
 intracellular accumulation of collagen VII, loss of anchoring

RT fibrils' and skin blistering";
 RL J. Biol. Chem. 273:19228-19234(1998).
 DR MGD; MGI:88452; Cx22al.
 DR InterPro; IPR00087; Collagen.
 DR InterPro; IPR00885; Fib.collagen_C.
 DR InterPro; IPR01007; VWFC.
 DR Pfam; PF01391; Collagen; 17.
 DR Pfam; PF0093; vwc; 1.
 DR ProDom; PDO02018; Fib.collagen_C; 1.
 DR SMART; SM0038; COLEI; 1.
 DR SMART; SM0214; vWC; 1.
 DR PROSITE; PS01248; VWFC; 1.
 DR KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 KW Glycoprotein; Collagen; Cartilage; Signal; Alternative splicing.
 FT (BY SIMILARITY); COLLAGEN ALPHA I(II) CHAIN.
 FT AMINO-TERMINAL PROPEPTIDE
 FT SIGNAL 1 25 POTENTIAL.
 FT CHAIN 154 1213
 FT PROPEP 1214 1459
 FT PROPEP 26 153
 FT DOMAIN 32 89
 FT DOMAIN 173 1186
 FT DOMAIN 1187 1213
 FT VARSPLIC 29 29
 FT VARSPLIC 30 98
 FT MISSING (IN SHORT ISOFORM).
 SQ SEQUENCE 1459 AA; 139154 MW; FCG84FAY7C532E7F2 CRC64;

RESULT 9

Qy 1 PAGPMPNGKDGKV 15
 ID CA12_MOUSE STANDARD; PRT; 1459 AA.
 AC P28481;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-DEC-1992 (Rel. 24, Last annotation update)
 DE Collagen alpha 1(II) chain precursor [Contains: Chondrocalcin].
 GN COL2A1.

Mus musculus (Mouse)

Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognath; Muridae; Murinae; Mus.
 OC NCBI_TaxId=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 MEDLINE=91358489; PubMed=188613;
 RA Mettsaranta M., Toman D., de Crombrugge B., Vuorio E.;
 RT "Mouse type II collagen gene. Complete nucleotide sequence, exon
 structure, and alternative splicing";
 RL J. Biol. Chem. 266:16662-16669(1991).
 RN [2]
 RP SEQUENCE OF 1455-1459 FROM N.A.
 MEDLINE=91274355; PubMed=2054384;
 RA Metsaranta M., Toman D., de Crombrugge B., Vuorio E.;
 RT "Specific hybridization probes for mouse type I, II, III and IX
 collagen mRNAs";
 RL Biochim. Biophys. Acta 1089:241-243(1991).
 CC -I- FUNCTION: COLLAGEN TYPE II IS SPECIFIC FOR CARTILAGINOUS TISSUES.
 CC -I- SUBUNIT: TIMERS OF IDENTICAL ALPHA 1(II) CHAINS.
 CC -I- PPTM: PROLINES AT THE THIRD POSITION OF THE TRIPETIDE REPEATING
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 CC -I- SIMILARITY: CONTAINS 1 WFRC DOMAIN.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See [http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch](http://www.isb-sib.ch/announce/or send an email to license@isb-sib.ch)).

DR EMBL; X57982; CAA41047.1; -.
 DR MGD; MGI:88452; Cx22al.
 DR InterPro; IPR00087; Collagen.
 DR InterPro; IPR00885; Fib.collagen_C.
 DR InterPro; IPR01007; VWFC.
 DR Pfam; PF01391; Collagen; 17.
 DR Pfam; PF0093; vwc; 1.
 DR ProDom; PDO02018; Fib.collagen_C; 1.
 DR SMART; SM0038; COLEI; 1.
 DR SMART; SM0214; vWC; 1.
 DR PROSITE; PS01248; VWFC; 1.
 DR KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 KW Glycoprotein; Collagen; Cartilage; Signal; Alternative splicing.
 FT (BY SIMILARITY); COLLAGEN ALPHA I(II) CHAIN.
 FT AMINO-TERMINAL PROPEPTIDE
 FT SIGNAL 1 25 POTENTIAL.
 FT CHAIN 154 1213
 FT PROPEP 1214 1459
 FT PROPEP 26 153
 FT DOMAIN 32 89
 FT DOMAIN 173 1186
 FT DOMAIN 1187 1213
 FT VARSPLIC 29 29
 FT VARSPLIC 30 98
 FT MISSING (IN SHORT ISOFORM).
 SQ SEQUENCE 1459 AA; 139154 MW; FCG84FAY7C532E7F2 CRC64;

RESULT 10

Qy 1 PAGPMPNGKDGKV 15
 ID CA12_CHICK STANDARD; PRT; 369 AA.
 AC P02460;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-NOV-1988 (Rel. 09, Last sequence update)
 DE Collagen alpha 1(II) chain precursor (Fragment).
 GN COL2A1.

OS Gallus gallus (Chicken)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OC NCBI_TaxId=9031;
 RN [1]
 RP SEQUENCE OF 1-193 FROM N.A.
 MEDLINE=83306892; PubMed=3840018;
 RA Deak F., Airgraes W.S., Kiss I., Sparks K.J., Goetinck P.F.;
 RT "Primary structure of the telopeptide and a portion of the helical
 domain of chicken type II procollagen as determined by DNA sequence
 analysis";
 RL Biochem. J. 229:189-196(1985).
 RN [2]
 RP SEQUENCE OF 82-369 FROM N.A.
 MEDLINE=84239738; PubMed=6330084;
 RA Sandell L.J., Prentice H.L., Kravis D., Upholt W.B.;
 RT "Structure and sequence of the chicken type II procollagen gene.
 Characterization of the region encoding the carboxyl-terminal
 telopeptide and propeptide";
 RL J. Biol. Chem. 259:7826-7834(1984).
 RN [3]
 RP SEQUENCE OF 114-369 FROM N.A.
 RA Ninomiya Y., Showalter A.M., van der Rest M., Seidah N.G.,
 RA Christen M., Olsen B.R.;
 RT "Structure and sequence of the carboxyl propeptide of chicken type II procollagen
 determined by DNA and protein sequence analysis";
 RL Biochemistry 23:617-624(1984).

CC -!- FUNCTION: COLLAGEN TYPE II IS SPECIFIC FOR CARTILAGINOUS TISSUES.
 CC -!- SUBUNIT: TRIMERS OF IDENTICAL ALPHA 1(II) CHAINS.
 CC -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPETIDE REPEATING
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/or_send_an_email_to_license@isb-sib.ch).

CC EMBL: X02663; CAA26499_1; --.
 DR EMBL: L00063; AAB9967_1; --.
 DR EMBL: L00061; AAB9967_1; JOINED.
 DR EMBL: L00062; AAB9967_1; JOINED.
 PIR: A02860; CGCHC.
 DR InterPro: IPR00085; Collagen.
 DR InterPro: IPR00107; VWF.
 Pfam: PF01410; COLFI; 1.
 DR Pfam: PF01391; Collagen; 1.
 DR Prodrom: PD002078; Fib_collagen_C; 1.
 SMART: SM00038; COLFI; 1.
 DR PROSITE: PS01208; VWF; PARTIAL.
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 Glycoprotein; Cartilage; Collagen.
 FT NON_TER 1
 CHAIN <1 123
 PROPEP 124 369
 FT DOMAIN <1 96
 DOMAIN 97 123
 FT CARBOHYD 270 270
 FT DISULFID 275 320
 SQ SEQUENCE 369 AA: EF530625B08A3B0 CRC64:
 Query Match Score 56; DB 1; Length 369;
 Best Local Similarity 66.7%; Pred. No. 0.88;
 Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
 Qy 1 PAGPWGPNGKDGKVG 15
 Db 47 PP GPVGPGPSGDGS 61

RESULT 11

CA13_RAT
 ID CA13_RAT STANDARD; PRT; 636 AA.
 AC P13941; 070604;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Collagen alpha 1(III) chain (fragment).
 GN COL1A1.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Chordata; Craniata; Vertebrata; Buteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBL_TaxID-1016;
 RN [1] SEQUENCE FROM N.A.
 RX MEDLINE-9414571; Pubmed-8286415;
 RA Glumoff V., Maekelae J.K., Vuorio E.;
 RT "Cloning of cDNA for rat pro alpha 1(III) collagen mRNA. Different expression patterns of type I and type III collagen and fibronectin genes in experimental granulation tissue.",
 RT Biochim. Biophys. Acta 1217:41-48(1994).
 RN [2] SEQUENCE OF 73-636 FROM N.A.
 RP STRAIN-Sprague-Dawley; TISSUE-Fibroblast;
 RA Wurtz T., Ellerstorp C., Lundmark C., Christersson C.;
 RL Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.
 [3]

RP SEQUENCE OF 308-482 FROM N.A.
 RX MEDLINE-88296083; Pubmed-2456904;
 RA Franklin F.R., Hsu C.-Y.J., Meyers J.C., Lin E., Lyttle C.R., Kom B., Mohn K.;
 RT "Regulation of alpha 2(II), alpha 1(III), and alpha 2(V) collagen mRNAs by estrogen in the immature rat uterus.",
 RL DNA 7:347-354(1988).
 CC -!- FUNCTION: COLLAGEN TYPE III OCCURS IN MOST SOFT CONNECTIVE TISSUES
 CC -!- ALONG WITH TYPE I COLLAGEN.
 CC -!- SUBUNIT: TRIMERS OF IDENTICAL ALPHA 1(III) CHAINS. THE CHAINS ARE
 CC LINKED TO EACH OTHER BY INTERCHAIN DISULFIDE BONDS. TRIMERS ARE
 CC ALSO CROSS-LINKED VIA HYDROXYLINES.
 CC EMBL: X0369; CAA9832_1; --.
 DR EMBL: A005395; CAA06510_1; --.
 DR EMBL: M21354; AAA0942_1; --.
 DR PIR: A29905; A29905.
 DR InterPro: IPR00085; Collagen.
 DR InterPro: IPR00107; VWF.
 Pfam: PF01410; COLFI; 1.
 DR SMART: SM00038; COLFI; 1.
 DR PROSITE: PS01208; VWF; PARTIAL.
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 Glycoprotein; Cartilage; Collagen.
 FT NON_TER 1
 CHAIN <1 123
 PROPEP 124 369
 FT DOMAIN <1 96
 DOMAIN 97 123
 FT CARBOHYD 270 270
 FT DISULFID 275 320
 SQ SEQUENCE 369 AA: EF530625B08A3B0 CRC64:
 Query Match Score 56; DB 1; Length 369;
 Best Local Similarity 66.7%; Pred. No. 0.88;
 Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
 Qy 1 PAGPWGPNGKDGKVG 15
 Db 47 PP GPVGPGPSGDGS 61

RESULT 12

CA1B_MOUSE
 ID CA1B_MOUSE STANDARD; PRT; 1804 AA.
 AC Q61245; 0640V7;
 DT 01-Nov-1997 (Rel. 35, Created)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Collagen alpha 1(XI) chain precursor.
 GN COL1A1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBL_TaxID-10090;
 RN [1] SEQUENCE FROM N.A.

CC	use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC	-----
CC	EMBL: J04177; AAA51891.1; -.
DR	EMBL; L38956; AAA51891.1; -.
DR	PIR: A31795; A31795.
DR	MIM: 120280;
DR	InterPro: IPR000067; Collagen.
DR	InterPro: IPR000865; Fib_collagen_C.
DR	InterPro: IPR00191; laminin_G.
DR	InterPro: IPR03139; TSPN.
DR	Pfam: PF01410; COLF1; 1.
DR	Pfam: PF01391; Collagen; 18.
DR	Pfam: PF02210; TSPN; 1.
DR	ProDom: PD002078; Fib_collagen_C; 1.
DR	SMART: SM00338; COLF1; 1.
DR	SMART; SM00282; Lang; 1.
DR	SMART; SM00210; TSPN; 1.
KW	Glycoprotein; Collagen; Signal; Alternative splicing.
FT	SIGNAL 1 36 POTENTIAL.
FT	PROPEP 37 511 AMINO-TERMINAL PROPEPTIDE (POTENTIAL).
FT	CHAIN 512 1563 COLLAGEN ALPHA 1(III) CHAIN.
FT	PROPEP 1564 1806 CARBOXYL-TERMINAL PROPEPTIDE.
FT	DOMAIN 37 419 NONHELICAL REGION.
FT	DOMAIN 420 508 TRIPLE-HELICAL REGION (INTERRUPTED).
FT	DOMAIN 509 511 SHORT NONHETICAL SEGMENT.
FT	DOMAIN 512 528 TELEOPETIDE.
FT	DOMAIN 529 1542 TRIPLE-HELICAL REGION.
FT	DOMAIN 1543 1563 NONHETICAL REGION (C-TERMINAL).
FT	CARBOHYD 1640 1640 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	SITE 612 612 CROSSLINKING.
FT	SITE 1452 1452 CROSSLINKING.
FT	VARSPLTC 261 299 YAPDIDTYEGEAEK/KEAESVTEGPVTEIAQK -> KKKSNKRKKMVTAKSKERSKKFPKSEKFSSKK SYQSKAKLGKV (IN ISOFORMS WITH EXON II).
FT	SEQUENCE 1806 AA: 181137 MW: C27AD86180BCEDF5 CRC64;
Qy	Query Match, Best Local Similarity 60.9%; Score 55; DB 1; Length 1806; Matches 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
Db	1 PAGPWGPNKGDKGVG 15 1226 PRQPQGPNGADGQG 1240
RESULT 14	
CA13_CHICK	
ID CA13_CHICK STANDARD: PRT; 1262 AA.	
AC P12105; P77958; P79759; 090794; 092029;	
DT 01-OCT-1989 (Rel. 12, Created)	
DT 16-OCT-2001 (Rel. 40, Last sequence update)	
DT 01-MAR-2002 (Rel. 41, Last annotation update)	
DE Collagen alpha 1(III) chain precursor (Fragments).	
GN COLJAI.	
OS Gallus gallus (Chicken).	
OC Eukaryota; Chordata; Craniata; Vertebrata; Buteleostomi; Archosauaria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.	
OX NCBI_TAXID=9031;	
RN [1]	SEQUENCE OF 1-886 FROM N.A.
RC TISSUE-Kidney;	
RX MEDLINE-9426842; PubMed-8206952;	
RA Nah H.-D., Niu Z., Adams S.L.;	
RT "An alternative transcript of the chick type III collagen gene that does not encode type III collagen.";	
RL J. Biol. Chem. 269:16443-16448(1994).	
RN [2]	SEQUENCE OF 29-96; 332-397; 431-484; 503-535 AND 869-976 FROM N.A.
RP RX MEDLINE-84270696; PubMed-654770;	
RA Yamada Y., Liou G., Mudryj M., Obici S., de Crombrugge B.;	
RT "Conservation of the sizes for one but not another class of exons in two chick collagen genes.";	
RT RL Nature 310:333-337(1984).	
RN [3]	SEQUENCE OF 977-1262 FROM N.A.
RP RX MEDLINE-83220816; PubMed-6856474;	
RA Yamada Y., Kuhn K., de Crombrugge B.;	
RT "A conserved nucleotide sequence, coding for a segment of the C-propeptide, is found at the same location in different collagen genes.";	
RT RL Nucleic Acids Res. 11:2733-2744(1983).	
CC CC -I FUNCTION: COLLAGEN TYPE III OCCURS IN MOST SOFT CONNECTIVE TISSUES	
CC CC -I ALONG WITH TYPE I COLLAGEN.	
CC CC -I SUBUNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.	
CC CC -I SIMILARITY: CONTAINS 1 VWFC DOMAIN.	
FT FT -----	
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).	
CC CC -----	
DR DR EMBL: U07973; AAA83407.1; .	
DR DR EMBL; X00822; CAB52686.1; JOINED.	
DR DR EMBL; X00823; CAB52686.1; JOINED.	
DR DR EMBL; X00825; CAB25397.1; ALT_SEQ.	
DR DR EMBL; X00827; CAB25398.1; .	
DR DR EMBL; X00828; CAB25399.1; .	
DR DR EMBL; X00830; CAB25401.1; .	
DR DR EMBL; X00831; CAB25402.1; .	
DR DR EMBL; X02302; AAD15299.1; .	
DR DR EMBL; K02301; AAD15298.1; .	
DR DR EMBL; M36662; AAA18519.1; ALT_SEQ.	
DR DR PIR: A05269; A05269.	
DR DR InterPro: IPR000867; Collagen.	
DR DR InterPro: IPR000865; Fib_collagen_C.	
DR DR InterPro: IPR01007; VWFC.	
DR DR Pfam: PF01391; Collagen; 13.	
DR DR ProDom: PD002078; Fib_collagen_C; 1.	
DR DR SMART; SM00214; VWFC; 1.	
DR DR PROSITE; PS01208; VWFC; 1.	
KW KW -----	
FT FT Extracellular matrix; Connective tissue; Repeat; Hydroxylation; Glycoprotein; Collagen; Signal; POTENTIAL.	
FT FT PROPEP 24 144 AMINO-TERMINAL PROPEPTIDE (BY SIMILARITY).	
FT FT CHAIN 145 1003 COLLAGEN ALPHA 1(III) CHAIN CARBOXYL-TERMINAL PROPEPTIDE (BY SIMILARITY).	
FT FT PROPEP 1004 1262 TRIPLE-HELICAL REGION (BY SIMILARITY).	
FT FT DOMAIN 229 88 VWFC.	
FT FT DOMAIN 145 164 NONHETICAL REGION (N-TERMINAL) (BY SIMILARITY).	
FT FT DOMAIN 165 994 INTERCHAIN (BY SIMILARITY).	
FT FT DOMAIN 995 995 TRIPLE-HETICAL REGION (BY SIMILARITY).	
FT FT DOMAIN 995 1003 NONHETICAL REGION (C-TERMINAL) (BY SIMILARITY).	
FT FT NON_CONS 886 887 INTERCHAIN (BY SIMILARITY).	
FT FT DISULFID 994 994 INTERCHAIN (BY SIMILARITY).	
FT FT DISULFID 995 995 HYDROXYLATION (BY SIMILARITY).	
FT FT MOD_RES 262 262 HYDROXYLATION (BY SIMILARITY).	
FT FT MOD_RES 283 283 HYDROXYLATION (BY SIMILARITY).	

MOD_RES 859 859 HYDROXYLATION (BY SIMILARITY).
 FT CARBOHYD 1163 1163 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 96 96 E -> K (IN REF. 2).
 FT CONFLICT 1132 1132 F -> S (IN REF. 3).
 FT SEQUENCE 1262 AA; 121249 MW; 96ABE7B2E9DEB43D CRC64;
 SQ [8]

Query Match 59 %; Score 55; DB 1; Length 1262;
 Best Local Similarity 66.7%; Pred. No. 3.7; 1; Mismatches 4; Indels 0; Gaps 0;
 Matches 10; Conservative 1; Mismatches 1; Mismatches 1; Indels 0; Gaps 0;

Oy 1 PAGPWGPNKGDKGVG 15
 Db 942 PPGPAGPPGKDGREG 956

RESULT 15
 CA21_CHICK CA21_CHICK STANDARD PRT; 1362 AA.
 AC P02467; 090795; 097979; P87422; Q90758; Q92014; P87491; Q90792;
 DT 21-JUL-1996 (Rel. 01, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Collagen alpha 2(I) chain precursor (Fragments).
 GN COLLA2.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Phasianinae; Gallus.
 OC NCBI_TAXID=9031;
 RN [1]
 RP SEQUENCE OF 1-245; 262-448 AND 466-1362 FROM N.A.
 RX MEDLINE=61185168; PubMed=3869961;
 RA Boedtker H., Finer M., Aho S.;
 RT "The structure of the chicken alpha 2 collagen gene.";
 RL Ann. N.Y. Acad. Sci. 460:85-116(1985).
 RN [2]
 RP SEQUENCE OF 1-89 FROM N.A.
 RX MEDLINE=61346518; PubMed=6135195;
 RA Tate V.E., Finer M.H., Boedtker H., Doty P.;
 RT "Chick pro alpha 2 (I) collagen gene: exon location and coding potential for the prepropeptide.";
 RL Nucleic Acids Res. 11:91-104(1983).
 RN [3]
 RP SEQUENCE OF 1-14 FROM N.A.
 RX MEDLINE=2206040; PubMed=6946474;
 RA Vogeli G., Ohkubo H., Sobel M.E., Yamada Y., Pastan I.,
 RT de Crombrugge B.;
 RT "Structure of the promoter for chicken alpha 2 type I collagen gene.";
 RL Proc. Natl. Acad. Sci. U.S.A. 78:5334-5338(1981).
 RN [4]
 RP SEQUENCE OF 1-33 FROM N.A.
 RX MEDLINE=64297217; PubMed=6473103;
 RA Aho S., Tate V.E., Boedtker H.;
 RT "Location of the 11 bp exon in the chicken pro alpha 2(I) collagen gene.";
 RL Nucleic Acids Res. 12:6117-6125(1984).
 RN [5]
 RP SEQUENCE OF 1-79 FROM N.A.
 RX MEDLINE=368056316; PubMed=3678834;
 RA Finer M.H., Boedtker H., Doty P.;
 RT "Construction and characterization of cDNA clones encoding the 5' end of the chicken pro alpha 1(I) collagen mRNA.";
 RL Gene 56:71-78(1987).
 RN [6]
 RP SEQUENCE OF 78-92.
 TISSUE-SKIN;
 RC MEDLINE=1115216; PubMed=5544653;
 RA Highbarger J.H., Kang A.H., Gross J.;
 RT "Comparative studies on the amino acid sequence of the alpha 2-CB2 peptides from chick and rat skin collagens.";
 RL Biochemistry 10:610-616(1971).
 RN [7]
 SEQUENCE OF 74-91; 263-448 AND 1088-1169 FROM N.A.

RX MEDLINE=82050801; PubMed=6272119;
 RA Wozney J., Hanahan D., Tate V.E., Boedtker H., Doty P.,
 RT "Structure of the pro alpha 2 (I) collagen gene.";
 RL Nature 294:129-135(1981).
 SQ [8]

RN SEQUENCE OF 78-92.
 RC TISSUE=Skin;
 RX MEDLINE=70131186; PubMed=4313735;
 RA Kang A.H., Gross J.;
 RT "Amino acid sequence of cyanogen bromide peptides from the amino-terminal region of chick skicollagen.";
 RL Biochemistry 9:796-804(1970).
 RN [9]
 RP SEQUENCE OF 78-92 AND 415-448.
 RC TISSUE=Skin;
 RX MEDLINE=69285369; PubMed=5809220;
 RA Kang A.H., Igashiki S., Gross J.;
 RT "Characterization of the peptides derived from the alpha-2 chain of chick skin collagen.";
 RL Biochemistry 8:3200-3204(1969).
 RN [10]
 RP SEQUENCE OF 78-92 AND 415-448.
 RC TISSUE=Bone;
 RX MEDLINE=6926882; PubMed=5785233;
 RA Lane J.M., Miller E.J.;
 RT "Isolation and characterization of the peptides derived from the alpha 2 chain of chick bone collagen after cyanogen bromide cleavage.";
 RL Biochemistry 8:2134-2139(1969).
 RN [11]
 RP SEQUENCE OF 566-587 FROM N.A.
 RX MEDLINE=9074829; PubMed=364479;
 RA Lehrach H., Frischaut A.-M., Hanahan D., Wozney J., Fuller F.,
 RA Crkvenjakov R., Boedtker H., Doty P.;
 RT "Construction and characterization of a 2.5-kilobase procollagen clone.";
 RL Proc. Natl. Acad. Sci. U.S.A. 75:5417-5421(1978).
 RN [12]
 RP SEQUENCE OF 902-1362 FROM N.A.
 RX MEDLINE=81160715; PubMed=6927845;
 RA Fuller F., Boedtker H.;
 RT "Sequence determination and analysis of the 3' region of chicken pro-alpha 1(I) and pro-alpha 2(I) collagen messenger ribonucleic acids including the carboxy-terminal propeptide sequences.";
 RL Biochemistry 20:996-1006(1981).
 RN [13]
 RP SEQUENCE OF 998-1159 AND 1234-1362 FROM N.A.
 RX MEDLINE=81264246; PubMed=6267043;
 RA Dickson L.A., Ninomaya Y., Bernard M.P., Pesciotta D.M., Parsons J.,
 RA Green G., Eikenberry E.F., de Crombrugge B., Vogeli G., Pastan I.,
 RA Fletzek P.P., Olsen B.R.;
 RT "The exon/intron structure of the 3'-region of the pro alpha 2(I) collagen gene.";
 RL J. Biol. Chem. 256:8407-8415(1981).
 RN [14]
 RP SEQUENCE OF 932-954 AND 968-980 FROM N.A.
 RX MEDLINE=81064671; PubMed=6159982;
 RA Avvedimento V.E., Vogeli G., Yamada Y., Maizel J.V. Jr., Pastan I.,
 RA de Crombrugge B.;
 RT "Correlation between splicing sites within an intron and their sequence complementarity with u1 RNA.";
 RL Cell 21:689-696(1980).
 RN [15]
 RP SEQUENCE OF 1266-161; 467-517 AND 926-954 FROM N.A.
 RX MEDLINE=81112157; PubMed=7460017;
 RA Iran M., Pastan I., de Crombrugge B.;
 RA Yamada Y., Avvedimento V.E., Mudryj M., Ohkubo H., Vogeli G.,
 RT "The collagen gene: evidence for its evolutionary assembly by amplification of a DNA segment containing an exon of 54 bp.";
 RL Cell 22:887-892(1980).
 CC -1- FUNCTION: TYPE I COLLAGEN IS A MEMBER OF GROUP I COLLAGEN
 CC (-1- SUBUNIT: TRIMERS OF ONE ALPHA 2 (I) AND TWO ALPHA 1 (I) CHAINS.

RP RT Nucleic Acids Res. 12:1025-1038(1984).

RN RT the human type II procollagen gene.";

[9] RP SEQUENCE OF 35-167 FROM N.A.

RX MEDLINE=89233138; PubMed=2714801;

RX RA S.M.W., Benson-Chanda V., Visning H., Ramirez F.;

RT "Organization of the exons coding for pro alpha 1(II) collagen N"-

RT propodeptide confirms a distinct evolutionary history of this domain of

RT the fibrillar collagen genes.";

RN Genomics 4: 438-441(1989).

RN [10] RP REVIEW ON VARIANTS.

RX MEDLINE=91184577; PubMed=2010058;

RX RA Kivilanemi H., Tromp G., Prockop D.J.;

RT "Mutations in collagen genes: causes of rare and some common diseases

RT in humans";

RT FASEB J. 5:2052-2060(1991).

RN [11] RP REVIEW ON VARIANTS.

RX MEDLINE=92255955; PubMed=9101280;

RX RA Kivilanemi H., Tromp G., Prockop D.J.;

RT "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-

RT associated collagen (type IX), and network-forming collagen (type X)

RT cause a spectrum of diseases of bone, cartilage, and blood vessels.";

RL Hum. Mutat. 9:300-315(1997).

RN [12] RP VARIANT SER-1074.

RX MEDLINE=90036909; PubMed=2572591;

RX RA Visning H., D'Alessio M., Lee B., Ramirez F., Godfrey M.,

RA Hollister D. W.;

RT "Glycine to serine substitution in the triple helical domain of pro-

RT alpha 1 (III) collagen results in a lethal perinatal form of short-

RT limbed dwarfism.";

RT J. Biol. Chem. 264:18265-18267(1989).

RN [13] RP VARIANT SEDC 1095-GLY--TYR-1330 DEL.

RX MEDLINE=89266907; PubMed=2543071;

RX RA Lee B., Visning H., Ramirez F., Rogers D., Rimoain D.;

RT "Identification of the molecular defect in a family with

RT spondyloepiphyseal dysplasia.";

RL Science 244:978-980(1989).

RN [14] RP VARIANT OSTEOARTHRITIS CYS-650.

RX MEDLINE=90370826; PubMed=1975633;

RX RA Alla-Kokko L., Baldwin C.T., Moskowitz R.W., Prockop D.J.;

RA RT "Single base mutation in the type II procollagen gene (COL2A1) as a

RT cause of primary osteoarthritis associated with a mild

RT chondroplasia.";

RT Proc. Natl. Acad. Sci. U.S.A. 87:6555-6558(1990).

RN [15] RP VARIANT OT-I-IV VAL-717.

RX MEDLINE=91291136; PubMed=2064612;

RX RA Bateman J.F., Hannagan M., Cole W.G.;

RA RT "Characterization of a type I collagen alpha 2(I) glycine-586 to

RT valine substitution in osteogenesis imperfecta type IV. Detection of

RT the mutation and prenatal diagnosis by a chemical cleavage method.";

RL Biochem. J. 276:765-770(1991).

RN [16] RP VARIANT OSTEOARTHRITIS CYS-650.

RX MEDLINE=91086471; PubMed=1985108;

RX RA Eyre D.R., Weis M.A., Moskowitz R.W.;

RT "Cartilage expression of a type II collagen mutation in an inherited

RT form of osteoarthritis associated with a mild chondroplasia.";

RL J. Clin. Invest. 87:357-361(1991).

RN [17] RP VARIANT HYPOCHONDROGENESIS GLU-984.

RX MEDLINE=93054548; PubMed=1429602;

RX RA Bogaert R., Tiller G.E., Wies M.A., Gruber H.E., Rimoain D.L.,

RA Cohn D.H., Eyer D.R.;

RA "An amino acid substitution (Gly853-->Glu) in the collagen alpha

RT 1(II) chain produces hypochondrogenesis.";

RL J. Biol. Chem. 267:22522-22526(1992).

RN [18] RP VARIANT WS-II ASP-198.

RX MEDLINE=93304428; PubMed=8317498;

RX RA Koerkkoje J., Rivanen P., Haataja L., Kaaeraiaeinen H.,

RA Kivirikko K.I., Prockop D.J., Ala-Kokko L.;

RT "Mutation in type II procollagen (COL2A1) that substitutes aspartate

RT for glycine alpha 1-67 and that causes cataracts and retinal

RT detachment: evidence for molecular heterogeneity in the Wagner

RT syndrome and the Stickler syndrome (arthro-ophtalmopathy).";

RL Am. J. Hum. Genet. 53:A209-A209(1993).

RN [20] RP VARIANT SEMD CYS-840.

RX Miller G.E., Weis M.A., Lachman R.S., Cohn D.H., Rimoain D.L.,

RA Eyer D. R.;

RT "A dominant mutation in the type II collagen gene (COL2A1) produces

RT spondyloepimetaphysial dysplasia (SEMD), Strudwick type.";

RL Am. J. Hum. Genet. 53:A209-A209(1993).

RN [21] RP VARIANT OSTEOARTHRITIS CYS 595.

RX MEDLINE=9328819; PubMed=8407190;

RX Holderbaum D., Malemud C.J., Moskowitz R.W., Haqqi T.M.;

RT "Human cartilage from late stage familial osteoarthritis transcribes

RT type II collagen mRNA encoding a cysteine in position 519.";

RL Biochem. Biophys. Res. Commun. 192:1169-1174(1993).

RN [22] RP VARIANT SEMD ARG-285.

RX MEDLINE=93252400; PubMed=8486375;

RX RA Viikula M., Rivanen P., Vuorio A.F., Kaitila I., Ala-Kokko L.,

RA Peitonen L.;

RT "A mutation in the amino-terminal end of the triple helix of type II

RT collagen causing severe osteochondrodysplasia.";

RL Genomics 16:282-285(1993).

RN [23] RP VARIANT SEDC CYS-206.

RX MEDLINE=94063862; PubMed=8244341;

RX Williams C.J., Considine E.L., Knowlton R.G., Reginato A., Neumann G.,

RA Harrison D., Buxton P., Jimenez S.A., Prockop D.J.;

RT "Spondyloepiphyseal dysplasia and precocious osteoarthritis in a

RT family with an Arg975-->Cys mutation in the procollagen type II gene

RT (COL2A1)." ;

RL Hum. Genet. 92:499-505(1993).

RN [24] RP VARIANT SEDC CYS-920.

RX MEDLINE=93315508; PubMed=8325895;

RX Chan D., Taylor T.K.F., Cole W.G.;

RT "Characterization of an arginine 789 to cysteine substitution in

RT alpha 1 (II) collagen chains of a patient with spondyloepiphyseal

RT dysplasia.";

RL J. Biol. Chem. 268:15238-15245(1993).

RN [25] RP VARIANT SEDC SER-1128.

RX MEDLINE=93140139; PubMed=8423604;

RX Cole W.G., Hall R. K., Rogers J.-G.;

RT "The clinical features of spondyloepiphyseal dysplasia congenita

RT resulting from the substitution of glycine 997 by serine in the alpha

RT 1(II) chain of type II collagen.";

RL J. Med. Genet. 30:27-35(1993).

Query Match 59.8%; Score 55; DB 1; Length 1418;

Best Local Similarity 66.7%; Pred. No. 4.2;

Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 1 PACPWPWPNKGDKGV 15.

Dy 1096 PPQPGVPGSGKDGGANG 1110

RESULT 17
 CALL_CHICK STANDARD; PRN; 1453 AA.
 ID CALL_CHICK STANDARD; PRN; 1453 AA.
 AC P02457;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-OCT-1988 (Rel. 12, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Collagen alpha 1(I) chain precursor.
 GN C01A1.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Archosaria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_TAXID=9031;
 RN [1] SEQUENCE OF 1-153 FROM N.A.
 RX MEDLINE=88056316; Published=3678834;
 RA Finer M.H., Boedtke H., Doty P.;
 RT "Construction and characterization of cDNA clones encoding the 5' end
 of the chicken pro-alpha 1(I) collagen mRNA.",
 RL Gene 56:71-78(1987).
 RN [2] SEQUENCE OF 1-144 FROM N.A.
 RX MEDLINE=88007542; Published=2830966;
 RA Finer M.H., Aho S., Gerstenfeld L.C., Boedtke H., Doty P.;
 RT "Unusual DNA sequences located within the promoter region and the
 first intron of the chicken pro-alpha 1(I) collagen gene.",
 RL J. Biol. Chem. 262:13323-13332(1987).
 RN [3] SEQUENCE OF 152-187.
 RX MEDLINE=82231995; Published=7093229;
 RA Higginbotham J.H., Corbett C., Dixit S.N., Yu W., Seyer J.M.,
 RA Kong A.H., Gross J.;
 RT "Amino acid sequence of chick skin collagen alpha 1(I)-CB8 and the
 complete primary structure of the helical portion of the chick skin
 collagen alpha 1(I) chain.",
 RL Biochemistry 21:2048-2055(1982).
 RN [4] SEQUENCE OF 1200-1205.
 RX MEDLINE=72243016; Published=5047697;
 RA Eyre D.R., Gimlicher M.J.;
 RT "Evidence for a previously undetected sequence at the carboxyterminus
 of the alpha 1 chain of chicken bone collagen.",
 RL Biochem. Biophys. Res. Commun. 48:720-726(1972).
 RN [5] SEQUENCE OF 981-1453 FROM N.A.
 RX MEDLINE=81160715; Published=6927845;
 RA Fuller F., Boedtke H.;
 RT "Sequence determination and analysis of the 3' region of chicken pro-
 alpha 1(I) and pro-alpha 2(I) collagen messenger ribonucleic acids
 including the carboxy-terminal propeptide sequences.",
 RL Biochemistry 20:996-1006(1981).
 RN [6] SEQUENCE OF 1311-1453 FROM N.A.
 RX MEDLINE=80134546; Published=6907008;
 RA Showalter A.M., Pesciotta D.M., Eikenberry E.F., Yamamoto T.,
 RA Pastan I., Decrombrughe B., Fietzke P.P., Olsen B.R.;
 RT "Nucleotide sequence of a collagen cDNA-fragment coding for the
 carboxyl end of pro alpha 1(I)-chains.",
 RL FEBS Lett. 111:61-65(1980).
 CC -I- FUNCTION: TYPE I COLLAGEN IS A MEMBER OF GROUP I COLLAGEN
 CC -I- (FIBRILLAR FORMING COLLAGEN).
 CC -I- SUBUNIT: TRIMERS OF ONE ALPHA 2(I) AND TWO ALPHA 1(I) CHAINS.
 CC -I- TISSUE SPECIFICITY: FORMS THE FIBRILS OF TENDON, LIGAMENTS AND
 CC BONES. IN BONES THE FIBRILS ARE MINERALIZED WITH CALCIUM
 CC HYDROXYAPATITE.
 CC -I- PROLINES AT THE THIRD POSITION OF THE TRIPETIDE REPEATING
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 CC -I- SIMILARITY: CONTAINS 1 WFPC DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M17839; AAA48704.1; JOINED.
 DR EMBL; M17838; AAA48704.1; JOINED.
 DR EMBL; V00401; CAM23695.1; -.
 DR EMBL; M10571; AAA8671.1; ALT_SEQ.
 DR EMBL; M17607; AAA48672.1; -.
 DR PIR; A02857; CGCHIS.
 DR PIR; A2179; A27179.
 DR A23367; A29367.
 DR InterPro; IPR000087; Collagen.
 DR InterPro; IPR000885; Fib_colagen_C.
 DR InterPro; IPR001007; WFCC.
 DR Pfam; PF01391; Collagen; 18.
 DR Pfam; PF00093; vwc; 1.
 DR Prodrom; PD002078; Fib_colagen_C; 1.
 DR SMART; SM00338; COLFI; 1.
 DR SMART; SM02124; VWC; 1.
 DR PROSITE; PS01208; WFCC; 1.
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 KW Glycoprotein; Collagen; Signal.
 FT SIGNAL 1 22
 FT PROPEP 23 151
 FT CHAIN 152 1205
 FT PROPEP 1 453
 FT DOMAIN 331 89
 FT MOD_RES 152 152
 FT PROTEIN 254 254
 FT MOD_RES 851 851
 FT MOD_RES 1081 1081
 FT MOD_RES 1097 1097
 FT MOD_RES 1153 1153
 FT CONFLICT 1187 1187
 FT CONFLICT 1441 1441
 FT CONFLICT 1441 1441
 SQ SEQUENCE 1453 AA; 137789 MW; 3BC61521342714D CRC64;
 Query Match 59.8%; Score 55; DB 1; Length 1453;
 Best Local Similarity 66.7%; Pred. No. 4.2;
 Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
 QY 1 PACPWPGRPNKGDKGVG 15
 DB 211 PAGPPGKNGDDGERG 225

RESULT 18
 CA13_MOUSE STANDARD; PRN; 1464 AA.
 ID CA13_MOUSE STANDARD; PRN; 1464 AA.
 AC P00121; Q61429; Q9CRN7; Created)
 DT 01-AUG-1988 (Rel. 08, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Collagen alpha 1(III) chain precursor.
 GN C01A1.
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL X DBA; TISSUE=Embryo;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBI_TAXID=10090;
 RN [1] SEQUENCE FROM N.A.
 RX MEDLINE=95011609; PubMed=7956795;
 RA Tomai D., de Crombrughe B.;
 RT "The mouse type-III procollagen-encoding gene: genomic cloning and
 RT complete DNA sequence.",
 RL Gene 147:161-168(1994).

RN [2] SEQUENCE OF 1-488 FROM N.A.
 RP RR InterPro: IPRO00087; Collagen.
 RX MEDLINE=88167858; PubMed=3443309;
 RA InterPro: IPRO00885; Fib_Collagen_C.
 RT Wood L., Theriault N., Vuoreli G.;
 "Complete nucleotide sequence of the N-terminal domains of the murine
 alpha-1 type-III collagen chain.";
 DR alpha-1 type-III collagen chain.; Gene 61:225-230(1987).
 RN [3] SEQUENCE OF 1-128 FROM N.A.
 RP RX MEDLINE=85131189; PubMed=3972847;
 RA Liu G., Mudry J.M., de Crombrugge B.;
 "Identification of the promoter and first exon of the mouse alpha 1
 (III) collagen gene.";
 RT (III) collagen gene.; J. Biol. Chem. 260:3773-3777(1985).
 RL DR SMART; SNO0214; VWA; 1.
 RP DR PROSITE; PS01208; VWFC; 1.
 RX KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 RC GLYCOPROTEIN; Collagen; Signal.
 RT FT SIGNAL 1 BY SIMILARITY.
 RT PROPEP 24 154 AMINO-TERMINAL PROPEPTIDE.
 FT CHAIN 155 1203 COLLAGEN ALPHA 1(III) CHAIN.
 FT PROPEP 1204 1464 CARBOXYL-TERMINAL PROPEPTIDE.
 FT DOMAIN 31 90 WFAC.
 FT DOMAIN 155 169 NONRHETICAL REGION (N-TERMINAL).
 FT DOMAIN 170 1195 TRIPLE-HELICAL REGION.
 FT DOMAIN 1196 1464 NONRHETICAL REGION (C-TERMINAL).
 FT CARBOHYD 262 262 O-LINKED (GAL- . .) (BY SIMILARITY).
 FT MOD_RES 262 262 HYDROXYLATION (BY SIMILARITY).
 FT MOD_RES 283 283 HYDROXYLATION (BY SIMILARITY).
 RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 Akizawa T., Hara D., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamakawa T.,
 Saito T., Okazaki Y., Gojobori T., Bonn H., Kasukawa T., Saito R.,
 Kadota K., Matsuda H.A., Ashburner M., Baralov S., Casavant T.,
 Fleischmann K., Gaasterland T., Gissi C., King B., Kochiwa H.,
 Kuehl P., Lewis S., Matsuoka Y., Nikaido I., Pesole G., Quackenbush J.,
 Schirali L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 Sakai K., Okido T., Furuno M., Baldarelli R., Barsh G.,
 Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 Gustincich S., Hill D., Hoffmann M., Hume D.A., Kamiya M., Lee N.H.,
 Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
 Wysniaowski A., Yoshida K., Hasegawa Y., Kawaiji H., Kohtsuki S.,
 RA "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 RN [5] SEQUENCE OF 1442-1464 FROM N.A.
 RP STRAIN=C57BL;
 RC MEDLINE=91274355; PubMed=2054304;
 RA Metsaaranta M., Toman D., de Crombrugge B., Vuorio E.;
 RT "Specific hybridization probes for mouse type I, II, III and IX
 collagen mRNAs";
 RL Biochim. Biophys. Acta 1089:241-243(1991).
 CC -I- FUNCTION: COLLAGEN TYPE III OCCURS IN MOST SOFT CONNECTIVE TISSUES
 -I- ALONG WITH TYPE I COLLAGEN.
 CC -I- SUBUNIT: TRIMERS OF IDENTICAL ALPHA 1(III) CHAINS. THE CHAINS ARE
 LINKED TO EACH OTHER BY INTRACHAIN DISULFIDE BONDS. TRIMERS ARE
 ALSO CROSS-LINKED VIA HYDROXYLINES.
 -I- PROLINES AT THE THIRD POSITION OF THE TRIPETIDE REPEATING
 UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 CC O-LINKED GLYCAN CONISTS OF GLC-GAL DISACCHARIDE (BY SIMILARITY).
 CC -I- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X57983; CAA36279.1; -.
 DR EMBL: M18933; AAA37338.1; -.
 DR EMBL: K03037; -, NOT_ANNOTATED_CDS.
 DR EMBL: AK019448; BAB31724.1; -.
 DR EMBL: X57983; CAA41048.1; -.
 DR PIR; A22287; A22287.
 DR PIR; A27353; A27353.
 DR PIR; S16373; S16373.
 DR MGI; B00453; Col3a1.
 RN [6] SEQUENCE OF 1-1592 FROM N.A.
 RP STRAIN=FVB/N; AND 129SV; TISSUE=cartilage;
 RX MEDLINE=97135795; PubMed=8901332;
 RA Vaudenbergh P., Vuoristo M.M., Ala-Kokko L., Prockop D.J.;
 RT "The mouse colla2 gene. Some transcripts from the adjacent rxr-beta
 gene extend into the colla2 gene.";
 RT Matrix Biol. 15:359-367(1996).
 CC -I- FUNCTION: MAY PLAY AN IMPORTANT ROLE IN FIBRILLGENESIS BY
 CONTROLLING LATERAL GROWTH OF COLLAGEN II FIBRILS (BY SIMILARITY).
 CC -I- SUBUNIT: TRIMERS COMPOSED OF THREE DIFFERENT CHAINS: ALPHA 1(XI),
 CC ALPHA 2(XI), AND ALPHA 3(XI). ALPHA 3(XI) IS A POST-TRANSLATIONAL
 CC MODIFICATION OF ALPHA 1(XI). ALPHA 1(V) CAN ALSO BE FOUND INSTEAD

CC OF ALPHA 3(XI)-1(II) (BY SIMILARITY).
 CC -1 ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS ARE PRODUCED BY
 CC ALTERNATIVE SPLICING.
 CC -1 PTM: PROLINES AT THE THIRD POSITION OF THE TRIPETIDE REPEATING
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 CC -1 SIMILARITY: BELONGS TO THE FIBRILLAR CLASS OF COLLAGENS.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement. (See <http://www.isb-sib.ch/announce/> or send an email to license@lsb-sib.ch).

CC

DR EMBL: AFI00956; AAC69905.1; - .

DR EMBL: U16789; AAA67751.1; - .

DR EMBL: U16790; AAA67752.1; - .

DR MGD; MGI:88447; COL1A2.

DR InterPro: IPR000087; Collagen.

DR InterPro: IPR000885; Fib collagen_C.

DR InterPro: IPR01791; Laminin_G.

DR InterPro: IPR003129; TSPN.

DR Pfam: PF01410; Colf1; 2.

DR Pfam: PF01391; Collagen; 18.

DR Pfam: PF02210; TSPN; 1.

DR Prodom: PDO0278; Fib collagen_C.

DR SMART: SM0038; Colf1; 1.

DR SMART: SM0282; Lamc; 1.

KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;

KW Glycoprotein; Collagen; Alternative splicing; Signal.

FT SIGNAL 1 22 POTENTIAL

FT CHAIN 23 1414 COLLAGEN ALPHA 2(XI) CHAIN.

FT PROPER 1415 1650 CARBOYL-TERMINAL PROTEPTIDE.

FT DOMAIN 23 255 NONHETICAL REGION.

FT DOMAIN 487 1414 TRIPLE-HELICAL REGION.

FT DOMAIN 1415 1650 NONHETICAL REGION.

FT CONFLICT 450 450 NO -> KP (IN REF. 2).

FT CONFLICT 618 619 V -> A (IN REF. 2).

FT CONFLICT 711 711 TGP -> HGS (IN REF. 2).

FT CONFLICT 757 759 A -> S (IN REF. 2).

FT CONFLICT 768 768 R -> G (IN REF. 2).

FT CONFLICT 790 790 R -> L (IN REF. 2).

FT CONFLICT 803 803 G -> V (IN REF. 2).

FT CONFLICT 835 836 G -> D (IN REF. 2).

FT CONFLICT 919 919 E -> V (IN REF. 2).

FT CONFLICT 1167 1167 P -> S (IN REF. 2).

FT CONFLICT 1300 1300 A -> T (IN REF. 2).

FT CONFLICT 1436 1436 I -> M (IN REF. 2).

SQ SEQUENCE 1650 AA; 162072 MW; 8ECDDBB702E71E2DA CRC64;

Query Match 59.8%; Score 55; DB 1; Length 1650;
 Best Local Similarity 66.7%; Prev. 4.8%;
 Matches 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 PAGPWPNGKDGKV 15
 Db 1098 PRGPAGPNGADGPG 112

RESULT 20

C2B_HUMAN STANDARD: PRT: 1736 AA. SEQUENCE FROM N.A.
 ID CA2B_HUMAN PRT: 1736 AA. SEQUENCE FROM N.A.
 AC P13942; Q13273; Q13271; Q13272; Q07751; Q99866; Q9UIP9;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Collagen alpha 2(XI) chain precursor.
 GN COL1A2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

NCBI_TAXID=9606; OX [1]

RN SEQUENCE FROM N.A.
 RP MEDLINE=96032717; PubMed=7559422;
 RX Vuoristo M.M., Pihlajamaa T., Vandenberg P., Prockop D.J., Ala-Kokko L.; "The human COL1A2 gene structure indicates that the gene has not evolved with the genes for the major fibrillar collagens.", J. Biol. Chem. 270:22873-22881(1995).
 RN [2]

RN SEQUENCE FROM N.A.
 RP Tubby B.; Submitted (Octr-1998) to the EMBL/GenBank/DDBJ databases.
 RN [3]

RN SEQUENCE OF 59-807 FROM N.A.
 RC TISSUE=Cartilage;
 RX MEDLINE=93114796; PubMed=8325374;
 RA Zhidkova N.I., Breerton R.G., Mayne R.; Molecular cloning of PARP (proline/arginine-rich protein) from human cartilage and subsequent demonstration that PARP is a fragment of the NH2-terminal domain of the collagen alpha 2(XI) chain"; FEBS Lett. 326:75-28(1993).
 RL [4]

RN SEQUENCE OF 730-1690 FROM N.A.
 RX MEDLINE=8834065; PubMed=2760050;
 RA Kimura T., Cheah K.S.E., Chan S.D.H., Lui V.C.H., Mattai M.-G., van der Rest M., Ono K., Solomon E., Niinomi Y., Olson B.R.; "The human alpha 2(XI) collagen (COL1A2) chain. Molecular cloning of cDNA and genomic DNA reveals characteristics of a fibrillar collagen with differences in genomic organization.", J. Biol. Chem. 264:13910-13916(1989).
 RN [5]

RN SEQUENCE OF 1-537 FROM N.A.
 RX MEDLINE=96435918; PubMed=8838804;
 RA Lui V.C., Ng L.J., Sat E.W., Cheah K.S.; "The human alpha 2(XI) collagen gene (COL1A2): completion of coding information, identification of the promoter sequence, and precise localization within the major histocompatibility complex reveal overlap with the KE5 gene.", Genomics 32:401-412(1995).
 RL [6]

RN [7]

RN ALTERNATIVE SPlicing.
 RX MEDLINE=9523868; PubMed=7721876;
 RA Zhdikova N.I., Justice S.K., Mayne R.; "Alternative mRNA processing occurs in the variable region of the pro-alpha 1(XI) and pro-alpha 2(XI) collagen chains.", J. Biol. Chem. 270:9486-9493(1995).
 RT REVIEW ON VARIANTS.
 RX MEDLINE=91255959; PubMed=9101290;
 RA Kuivaniemi H., Tromp G., Prockop D.J.; "Mutations in fibrilla collagen (types I, II, III, and XI), and network-forming collagen (type X) associated collagen (type IX), and cause a spectrum of diseases of bone, cartilage, and blood vessels.", Hum. Mutat. 9:300-315(1997).
 RL [8]

RN VARIANT ARG-661.
 RX MEDLINE=95163096; PubMed=7859284;
 RA Vikkula M., Matiman E.C.M., Lui V.C.H., Zhdikova N.I., Tiller G.B., van den Hoogen F.H.J., Ropers H.-H., Mayne R., Cheah K.S.E., Olsen B.R., Warman M.L., Brunner H.G.; "Autosomal dominant and recessive osteochondrodysplasias associated with the COL1A2 locus.", Cell 80:431-437(1995).
 RN [9]

RN VARIANT ASN-808.
 RX MEDLINE=2004768; PubMed=10581026;
 RA McGuire W.T., Prasad S.D., Griffith A.J., Kunst H.P.M., Green G.E., Shparigel K.B., Runge C., Huybrechts C., Mueller R.F., Lynch E., King M.-C., Brunner H.G., Creemers C.W.R.J., Takatori M., Li S.-W., Arita M., Mayne R., Prockop D.J., Van Camp G., Smith R.J.H.; "Mutations in COL1A2 cause non-syndromic hearing loss (DFNA13).",

RL
 CC
 CC
 CC
 -1- FUNCTION: MAY PLAY AN IMPORTANT ROLE IN FIBRILlogenesis BY CONTROLLING LATERAL GROWTH OF COLLAGEN II FIBRILS.
 -1- SUBUNIT: TRIMERS COMPOSED OF THREE DIFFERENT CHAINS: ALPHA 1(XI), ALPHA 2(XI), AND ALPHA 3(XI). ALPHA 3(XI) IS A POST-TRANSLATIONAL MODIFICATION OF ALPHA 1(II). ALPHA 1(V) CAN ALSO BE FOUND INSTEAD OF ALPHA 3(XI)=1(II).
 -1- ALTERNATIVE PRODUCTS: ALTERNATIVE FORMS OF THIS COLLAGEN ARE PRODUCED BY ALTERNATIVE SPLICING. THESE FORMS DIFFER IN THE PRESENCE OR ABSENCE OF THREE SEPARATE REGIONS IN THE N-TERMINAL NON-COLLAGENOUS DOMAIN.
 -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRipeptIDE REPEATING UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 -1- PROTEIN: A DISULFIDE-BONDED PEPTIDE CALLED PROLINE/ARGININE-RICH PROTEIN OR PARP IS RELEASED FROM THE AMINO TERMINUS DURING EXTRACELLULAR PROCESSING AND IS SUBSEQUENTLY RETAINED IN THE CARTILAGE MATRIX FROM WHICH IT CAN BE ISOLATED IN SIGNIFICANT AMOUNTS.

-1- DISEASE: DEFECTS IN COL1A2 ARE THE CAUSE OF AUTOSOMAL DOMINANT AND RECESSIVE OSTEOCHONDROPLASIS. (1) THE AUTOSOMAL DOMINANT FORM OF STICKLER SYNDROME (SS) IS CHARACTERIZED BY MILD SPONDYLOEPIPHYSEAL DYSPLASIA, OSTEOARTHRITIS, AND SENSORINEURAL HEARING LOSS. (2) THE AUTOSOMAL RECESSIVE DISORDER IS SIMILAR BUT MORE SEVERE.

-1- DISEASE: DEFECTS IN COL1A2 ARE THE CAUSE OF AUTOSOMAL DOMINANT NONSYNDROMIC SENSORINEURAL DEAFNESS TYPE 13 (DFNA13). AFFECTED INDIVIDUALS EXPERIENCE PROGRESSIVE HEARING LOSS BEGINNING IN THE SECOND TO FOURTH DECADES, EVENTUALLY MAKING USE OF AMPLIFICATION MANDATORY.

-1- SIMILARITY: BELONGS TO THE FIBRILLAR CLASS OF COLLAGENS.

-1- DATABASE: NAME="Hereditary hearing Loss homepage;
 NOTE="Gene page;
 WWW="http://www.via.ac.be/analab/hhh/hhgenes.html".

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/).

FT	PROPEP	1501	1736	CARBOXYL-TERMINAL PROPEPTIDE.
FT	DOMAIN	23	255	NONHETERICAL REGION.
FT	DOMAIN	487	1500	TRIPLE HELICAL REGION.
FT	DOMAIN	1501	1736	NONHETERICAL REGION.
FT	DOMAIN	10	18	POLY-LEU.
FT	CARBOHYD	298	301	POLY-GLU.
FT	CARBOHYD	1604	1604	N-LINKED (GlcNAc- <i>n</i>) (POTENTIAL).
FT	VARSPLIC	267	292	MISSING (IN ISOFORMS WITHOUT EXON 5).
FT	VARSPLIC	293	313	MISSING (IN ISOFORMS WITHOUT EXON 7).
FT	VARSPLIC	314	373	G -> R (IN AUTOSOMAL RECESSIVE OSTEochondrodysplasia).
FT	VARIANT	661	661	/FTid=VAR_001907.
				G -> E (IN DFNA13).
				/FTid=VAR_010655.
				R -> C (IN DFNA13).
				/FTid=VAR_010656.
				C -> G (IN REF. 1).
				S -> P (IN REF. 3 AND 5).
				Q -> R (IN REF. 3 AND 5).
				SL -> PP (IN REF. 2).
				A -> P (IN REF. 3).
				MP -> TL (IN REF. 3).
				AQ -> PR (IN REF. 3).
				NO -> KP (IN REF. 3).
				R -> Q (IN REF. 3).
				D -> N (IN REF. 3).
				TGPR -> HGST (IN REF. 4).
				QGP -> SG5 (IN REF. 4).
				P -> L (IN REF. 3).
				PP -> RQ (IN REF. 1 AND 4).
				D -> V (IN REF. 4).
				A -> R (IN REF. 4).
				ERGARGP -> GAGGGT (IN REF. 4).
				A -> P (IN REF. 2).
				Q -> T (IN REF. 2).
				E -> R (IN REF. 4).
FT	CONFLICT	894	894	P -> L (IN REF. 1).
FT	CONFLICT	704	705	PP -> RQ (IN REF. 1 AND 4).
FT	CONFLICT	97	97	D -> V (IN REF. 4).
FT	CONFLICT	530	531	A -> R (IN REF. 4).
FT	CONFLICT	531	531	ERGARGP -> GAGGGT (IN REF. 4).
FT	CONFLICT	726	726	A -> P (IN REF. 2).
FT	CONFLICT	542	542	Q -> T (IN REF. 2).
FT	CONFLICT	548	549	E -> R (IN REF. 4).
FT	CONFLICT	578	579	P -> L (IN REF. 1).
FT	CONFLICT	720	720	PP -> RQ (IN REF. 1 AND 4).
FT	CONFLICT	726	726	D -> V (IN REF. 4).
FT	CONFLICT	726	726	A -> R (IN REF. 4).
FT	CONFLICT	843	846	ERGARGP -> GAGGGT (IN REF. 4).
FT	CONFLICT	842	884	A -> P (IN REF. 2).
FT	CONFLICT	894	894	Q -> T (IN REF. 2).
FT	CONFLICT	1031	1032	E -> R (IN REF. 4).
FT	CONFLICT	1091	1091	P -> L (IN REF. 1).
FT	CONFLICT	1124	1124	PP -> RQ (IN REF. 1 AND 4).
FT	CONFLICT	1127	1133	D -> V (IN REF. 4).
FT	CONFLICT	1253	1253	A -> P (IN REF. 2).
FT	CONFLICT	1257	1257	Q -> T (IN REF. 2).
FT	CONFLICT	1552	1552	E -> R (IN REF. 4).
FT	CONFLICT	1736	AA;	PP -> RQ (IN REF. 1 AND 4).
FT	CONFLICT	1736	AA;	D -> V (IN REF. 4).
FT	CONFLICT	1736	AA;	A -> R (IN REF. 4).
FT	CONFLICT	1736	AA;	ERGARGP -> GAGGGT (IN REF. 4).
FT	CONFLICT	1736	AA;	A -> P (IN REF. 2).
FT	CONFLICT	1736	AA;	Q -> T (IN REF. 2).
FT	CONFLICT	1736	AA;	E -> R (IN REF. 4).
FT	CONFLICT	1736	AA;	P -> L (IN REF. 1).
FT	CONFLICT	1736	AA;	PP -> RQ (IN REF. 1 AND 4).
FT	CONFLICT	1736	AA;	D -> V (IN REF. 4).
FT	CONFLICT	1736	AA;	A -> R (IN REF. 4).
FT	CONFLICT	1736	AA;	ERGARGP -> GAGGGT (IN REF. 4).
FT	CONFLICT	1736	AA;	A -> P (IN REF. 2).
FT	CONFLICT	1736	AA;	Q -> T (IN REF. 2).
FT	CONFLICT	1736	AA;	E -> R (IN REF. 4).
FT	CONFLICT	1736	AA;	P -> L (IN REF. 1).
FT	CONFLICT	1736	AA;	PP -> RQ (IN REF. 1 AND 4).
FT	CONFLICT	1736	AA;	D -> V (IN REF. 4).
FT	CONFLICT	1736	AA;	A -> R (IN REF. 4).
FT	CONFLICT	1736	AA;	ERGARGP -> GAGGGT (IN REF. 4).
FT	CONFLICT	1736	AA;	A -> P (IN REF. 2).
FT	CONFLICT	1736	AA;	Q -> T (IN REF. 2).
FT	CONFLICT	1736	AA;	E -> R (IN REF. 4).
FT	CONFLICT	1736	AA;	P -> L (IN REF. 1).
FT	CONFLICT	1736	AA;	PP -> RQ (IN REF. 1 AND 4).
FT	CONFLICT	1736	AA;	D -> V (IN REF. 4).
FT	CONFLICT	1736	AA;	A -> R (IN REF. 4).
FT	CONFLICT	1736	AA;	ERGARGP -> GAGGGT (IN REF. 4).
FT	CONFLICT	1736	AA;	A -> P (IN REF. 2).
FT	CONFLICT	1736	AA;	Q -> T (IN REF. 2).
FT	CONFLICT	1736	AA;	E -> R (IN REF. 4).
FT	CONFLICT	1736	AA;	P -> L (IN REF. 1).
FT	CONFLICT	1736	AA;	PP -> RQ (IN REF. 1 AND 4).
FT	CONFLICT	1736	AA;	D -> V (IN REF. 4).
FT	CONFLICT	1736	AA;	A -> R (IN REF. 4).
FT	CONFLICT	1736	AA;	ERGARGP -> GAGGGT (IN REF. 4).
FT	CONFLICT	1736	AA;	A -> P (IN REF. 2).
FT	CONFLICT	1736	AA;	Q -> T (IN REF. 2).
FT	CONFLICT	1736	AA;	E -> R (IN REF. 4).
FT	CONFLICT	1736	AA;	P -> L (IN REF. 1).
FT	CONFLICT	1736	AA;	PP -> RQ (IN REF. 1 AND 4).
FT	CONFLICT	1736	AA;	D -> V (IN REF. 4).
FT	CONFLICT	1736	AA;	A -> R (IN REF. 4).
FT	CONFLICT	1736	AA;	ERGARGP -> GAGGGT (IN REF. 4).
FT	CONFLICT	1736	AA;	A -> P (IN REF. 2).
FT	CONFLICT	1736	AA;	Q -> T (IN REF. 2).
FT	CONFLICT	1736	AA;	E -> R (IN REF. 4).
FT	CONFLICT	1736	AA;	P -> L (IN REF. 1).
FT	CONFLICT	1736	AA;	PP -> RQ (IN REF. 1 AND 4).
FT	CONFLICT	1736	AA;	D -> V (IN REF. 4).
FT	CONFLICT	1736	AA;	A -> R (IN REF. 4).
FT	CONFLICT	1736	AA;	ERGARGP -> GAGGGT (IN REF. 4).
FT	CONFLICT	1736	AA;	A -> P (IN REF. 2).
FT	CONFLICT	1736	AA;	Q -> T (IN REF. 2).
FT	CONFLICT	1736	AA;	E -> R (IN REF. 4).
FT	CONFLICT	1736	AA;	P -> L (IN REF. 1).
FT	CONFLICT	1736	AA;	PP -> RQ (IN REF. 1 AND 4).
FT	CONFLICT	1736	AA;	D -> V (IN REF. 4).
FT	CONFLICT	1736	AA;	A -> R (IN REF. 4).
FT	CONFLICT	1736	AA;	ERGARGP -> GAGGGT (IN REF. 4).
FT	CONFLICT	1736	AA;	A -> P (IN REF. 2).
FT	CONFLICT	1736	AA;	Q -> T (IN REF. 2).
FT	CONFLICT	1736	AA;	E -> R (IN REF. 4).
FT	CONFLICT	1736	AA;	P -> L (IN REF. 1).
FT	CONFLICT	1736	AA;	PP -> RQ (IN REF. 1 AND 4).
FT	CONFLICT	1736	AA;	D -> V (IN REF. 4).
FT	CONFLICT	1736	AA;	A -> R (IN REF. 4).
FT	CONFLICT	1736	AA;	ERGARGP -> GAGGGT (IN REF. 4).
FT	CONFLICT	1736	AA;	A -> P (IN REF. 2).
FT	CONFLICT	1736	AA;	Q -> T (IN REF. 2).
FT	CONFLICT	1736	AA;	E -> R (IN REF. 4).
FT	CONFLICT	1736	AA;	P -> L (IN REF. 1).
FT	CONFLICT	1736	AA;	PP -> RQ (IN REF. 1 AND 4).
FT	CONFLICT	1736	AA;	D -> V (IN REF. 4).
FT	CONFLICT	1736	AA;	A -> R (IN REF. 4).
FT	CONFLICT	1736	AA;	ERGARGP -> GAGGGT (IN REF. 4).
FT	CONFLICT	1736	AA;	A -> P (IN REF. 2).
FT	CONFLICT	1736	AA;	Q -> T (IN REF. 2).
FT	CONFLICT	1736	AA;	E -> R (IN REF. 4).
FT	CONFLICT	1736	AA;	P -> L (IN REF. 1).
FT	CONFLICT	1736	AA;	PP -> RQ (IN REF. 1 AND 4).
FT	CONFLICT	1736	AA;	D -> V (IN REF. 4).
FT	CONFLICT	1736	AA;	A -> R (IN REF. 4).
FT	CONFLICT	1736	AA;	ERGARGP -> GAGGGT (IN REF. 4).
FT	CONFLICT	1736	AA;	A -> P (IN REF. 2).
FT	CONFLICT	1736	AA;	Q -> T (IN REF. 2).
FT	CONFLICT	1736	AA;	E -> R (IN REF. 4).
FT	CONFLICT	1736	AA;	P -> L (IN REF. 1).
FT	CONFLICT	1736	AA;	PP -> RQ (IN REF. 1 AND 4).
FT	CONFLICT	1736	AA;	D -> V (IN REF. 4).
FT	CONFLICT	1736	AA;	A -> R (IN REF. 4).
FT	CONFLICT	1736	AA;	ERGARGP -> GAGGGT (IN REF. 4).
FT	CONFLICT	1736	AA;	A -> P (IN REF. 2).
FT	CONFLICT	1736	AA;	Q -> T (IN REF. 2).
FT	CONFLICT	1736	AA;	E -> R (IN REF. 4).
FT	CONFLICT	1736	AA;	P -> L (IN REF. 1).
FT	CONFLICT	1736	AA;	PP -> RQ (IN REF. 1 AND 4).
FT	CONFLICT	1736	AA;	D -> V (IN REF. 4).
FT	CONFLICT	1736	AA;	A -> R (IN REF. 4).
FT	CONFLICT	1736	AA;	ERGARGP -> GAGGGT (IN REF. 4).
FT	CONFLICT	1736	AA;	A -> P (IN REF. 2).
FT	CONFLICT	1736	AA;	Q -> T (IN REF. 2).
FT	CONFLICT	1736	AA;	E -> R (IN REF. 4).
FT	CONFLICT	1736	AA;	P -> L (IN REF. 1).
FT	CONFLICT	1736	AA;	PP -> RQ (IN REF. 1 AND 4).
FT	CONFLICT	1736	AA;	D -> V (IN REF. 4).
FT	CONFLICT	1736	AA;	A -> R (IN REF. 4).
FT	CONFLICT	1736	AA;	ERGARGP -> GAGGGT (IN REF. 4).
FT	CONFLICT	1736	AA;	A -> P (IN REF. 2).
FT	CONFLICT	1736	AA;	Q -> T (IN REF. 2).
FT	CONFLICT	1736	AA;	E -> R (IN REF. 4).
FT	CONFLICT	1736	AA;	P -> L (IN REF. 1).
FT	CONFLICT	1736	AA;	PP -> RQ (IN REF. 1 AND 4).
FT	CONFLICT	1736	AA;	D -> V (IN REF. 4).
FT	CONFLICT	1736	AA;	A -> R (IN REF. 4).
FT	CONFLICT	1736	AA;	ERGARGP -> GAGGGT (IN REF. 4).
FT	CONFLICT	1736	AA;	A -> P (IN REF. 2).
FT	CONFLICT	1736	AA;	Q -> T (IN REF. 2).
FT	CONFLICT	1736	AA;	E -> R (IN REF. 4).
FT	CONFLICT	1736	AA;	P -> L (IN REF. 1).
FT	CONFLICT	1736	AA;	PP -> RQ (IN REF. 1 AND 4).
FT	CONFLICT	1736	AA;	D -> V (IN REF. 4).
FT	CONFLICT	1736	AA;	A -> R (IN REF. 4).
FT	CONFLICT	1736	AA;	ERGARGP -> GAGGGT (IN REF. 4).
FT	CONFLICT	1736	AA;	A -> P (IN REF. 2).
FT	CONFLICT	1736	AA;	Q -> T (IN REF. 2).
FT	CONFLICT	1736	AA;	E -> R (IN REF. 4).
FT	CONFLICT	1736	AA;	P -> L (IN REF. 1).
FT	CONFLICT	1736	AA;	PP -> RQ (IN REF. 1 AND 4).
FT	CONFLICT	1736	AA;	D -> V (IN REF. 4).
FT	CONFLICT	1736	AA;	A -> R (IN REF. 4).
FT	CONFLICT	1736	AA;	ERGARGP -> GAGGGT (IN REF. 4).
FT	CONFLICT	1736	AA;	A -> P (IN REF. 2).
FT	CONFLICT	1736	AA;	Q -> T (IN REF. 2).
FT	CONFLICT	1736	AA;	E -> R (IN REF. 4).
FT	CONFLICT	1736	AA;	P -> L (IN REF. 1).
FT	CONFLICT	1736	AA;	PP -> RQ (IN REF. 1 AND 4).
FT	CONFLICT	1736	AA;	D -> V (IN REF. 4).
FT	CONFLICT	1736	AA;	A -> R (IN REF. 4).
FT	CONFLICT	1736	AA;	ERGARGP -> GAGGGT (IN REF. 4).
FT	CONFLICT	1736	AA;	A -> P (IN REF. 2).
FT	CONFLICT	1736	AA;	Q -> T (IN REF. 2).
FT	CONFLICT	1736	AA;	E -> R (IN REF. 4).
FT	CONFLICT	1736	AA;	P -> L (IN REF. 1).
FT	CONFLICT	1736	AA;	PP -> RQ (IN REF. 1 AND 4).
FT	CONFLICT	1736	AA;	D -> V (IN REF. 4).
FT	CONFLICT	1736	AA;	A -> R (IN REF. 4).
FT	CONFLICT	1736	AA;	ERGARGP -> GAGGGT (IN REF. 4).
FT	CONFLICT	1736	AA;	A -> P (IN REF. 2).
FT	CONFLICT	1736	AA;	Q -> T (IN REF. 2).
FT	CONFLICT	1736	AA;	E -> R (IN REF. 4).
FT	CONFLICT	1736	AA;	P -> L (IN REF. 1).
FT	CONFLICT	1736	AA;	PP -> RQ (IN REF. 1 AND 4).
FT	CONFLICT	1736	AA;	D -> V (IN REF. 4).
FT	CONFLICT	1736	AA;	A -> R (IN REF. 4).
FT	CONFLICT			

RA Sulston J., Thlerry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
 RA Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,
 RA Wohldman P.; "2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans";
 RT Nature 368:32-38(1994).
 -1- FUNCTION: NEMATODE CUTICLES ARE COMPOSED LARGELY OF COLLAGEN-LIKE PROTEINS. THE CUTICLE FUNCTIONS BOTH AS AN EXOSKELETON AND AS A BARRIER TO PROTECT THE WORM FROM ITS ENVIRONMENT (BY SIMILARITY).
 -1- SUBUNIT: COLLAGEN POLYPEPTIDE CHAINS ARE COMPLEXED WITHIN THE CUTICLE BY DISULFIDE BONDS AND OTHER TYPES OF COVALENT CROSS-LINKS (BY SIMILARITY).
 CC -1- SIMILARITY: TO OTHER COLLAGENS. STRONG, TO OTHER CUTICLE COLLAGENS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; L23651; AAA27958 1; --.
 DR PIR: S44767; S44767.
 DR WormPep; C29E4_1; CE00083.
 DR InterPro; IPR002436; Col cuticle_N.
 DR InterPro; IPR000087; Collagen.
 DR Pfam; PF01391; Collagen; 2.
 DR Pfam; PF01484; Col cuticle_N; 1.
 KW Hypothetical protein; Cuticle; Connective tissue; Repeat;
 KW Multigene family; Collagen.
 FT DOMAIN 96 125 TRIPLE-HELICAL REGION.
 FT DOMAIN 142 204 TRIPLE-HELICAL REGION.
 FT DOMAIN 208 252 TRIPLE-HELICAL REGION.
 FT DOMAIN 256 270 TRIPLE-HELICAL REGION.

SQ SEQUENCE 305 AA; 30623 MW; BA4CCED5501E27D CRC64;

Query Match 58.7%; Score 54; DB 1; Length 305:
 Best Local Similarity 66.7%; Pred No. 1.4;
 Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Db 1 PAGGWPNGKDGKV 15
 111 11 : 1 11
 140 PAGPAGPPGQQGPVG 154

RESULT 22

CA13_BOVIN ID CA13_BOVIN STANDARD: PRT: 1049 AA.
 AC P04258;
 DT 20-MAR-1987 (Rel. 04, created)
 DT 20-MAR-1987 (Rel. 04, last sequence update)
 DT 16-OCT-2001 (Rel. 40, last annotation update)

DE Collagen alpha 1(III) chain.

GN COL3A1.

OS Bos taurus (bovine).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bovinae; Bos.

OC NCBI_TAXID=9913;
 RN [1]
 SEQUENCE OF 1-242.

RX MEDLINE=80026026; PubMed=488906;

RA Fietzek P.P., Altmann H., Rauterberg J., Henkel W., Wachter E., Kuhn K.; "The covalent structure of calf skin type III collagen. I-The amino acid sequence of the amino terminal region of the alpha 1(III) chain (positions 1-222)", Hoppe-Seyler's Z. Physiol. Chem. 360:809-820(1979).

RX SEQUENCE OF 243-422.

RX MEDLINE=80026027; PubMed=488907;

RA Dewes H., Fietzek P.P., Kuhn K.; "The covalent structure of calf skin type III collagen. II. The amino acid sequence of the cyanogen bromide peptide alpha 1(III)CB1,8,10,2 (positions 223-402);", Hoppe-Seyler's Z. Physiol. Chem. 360:821-832(1979).
 RN [3]
 SEQUENCE OF 423-571.
 RX MEDLINE=80026028; PubMed=488908;
 RA Benz H., Fietzek P.P., Kuhn K.; "The covalent structure of calf skin type III collagen. III. The amino acid sequence of the cyanogen bromide peptide alpha 1(III)CB4 (positions 403-551);", Hoppe-Seyler's Z. Physiol. Chem. 360:833-840(1979).
 RN [4]
 SEQUENCE OF 572-808.
 RX MEDLINE=80026029; PubMed=488909;
 RA Lang H., Gianville R.W., Fietzek P.P., Kuhn K.; "The amino acid sequence of the cyanogen bromide peptide alpha 1(III)CB5 (positions 555-788);", Hoppe-Seyler's Z. Physiol. Chem. 360:841-850(1979).
 RN [5]
 SEQUENCE OF 809-947.
 RX MEDLINE=80026030; PubMed=488910;
 RA Dewes H., Fietzek P.P., Kuhn K.; "The covalent structure of calf skin type III collagen. V. The amino acid sequence of the cyanogen bromide peptide alpha 1(III)CB9A (position 789-927);", Hoppe-Seyler's Z. Physiol. Chem. 360:851-860(1979).
 RN [6]
 SEQUENCE OF 948-1049.
 RX MEDLINE=80026031; PubMed=488911;
 RA Altmann H., Fietzek P.P., Gianville R.W., Kuhn K.; "The covalent structure of calf skin type III collagen. VI. The amino acid sequence of the carboxy-terminal cyanogen bromide peptide alpha 1(III)CB9B (positions 928-1028);", Hoppe-Seyler's Z. Physiol. Chem. 360:861-868(1979).
 -1- FUNCTION: COLLAGEN TYPE III OCCURS IN MOST SOFT CONNECTIVE TISSUES ALONG WITH TYPE I COLLAGEN.
 -1- SUBUNIT: TRIMERS OF IDENTICAL ALPHA 1(III) CHAINS. THE CHAINS ARE LINKED TO EACH OTHER BY INTERCHAIN DISULFIDE BONDS. TRIMERS ARE ALSO CROSS-LINKED VIA HYDROXYLYSINES.
 CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPÉPIPE DIPEPTIDE REPEATING UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 DR InterPro; IPR000087; Collagen.
 DR InterPro; IPR01007; VWC.
 DR Pfam; PF01391; Collagen; 17.
 DR PROSITE; PS01208; VWF_C; PARTIAL.
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 KW Glycoprotein; Collagen.
 FT DOMAIN 1 14 NONHETICAL REGION (N-TERMINAL).
 FT DOMAIN 15 1040 TRIPLE-HELICAL REGION.
 FT DOMAIN 1041 1049 NONHETICAL REGION (C-TERMINAL).
 FT MOD_RES 95 95 HYDROXYLATION.
 FT MOD_RES 107 107 HYDROXYLATION.
 FT MOD_RES 119 119 HYDROXYLATION.
 FT MOD_RES 938 938 HYDROXYLATION.
 FT MOD_RES 950 950 HYDROXYLATION.
 FT CARBOHYD 107 107 O-LINKED (GAL, . .).
 FT DISULFDID 1040 1040 INTERCHAIN.
 FT DISULFDID 1041 1041 INTERCHAIN.
 SQ SEQUENCE 1049 AA; 93651 MW; 8BECC33DIC66EC9A3 CRC64;

Query Match 58.7%; Score 54; DB 1; Length 1049:
 Best Local Similarity 66.7%; Pred No. 4.3;
 Matches 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Db 1 PAGGWPNGKDGKV 15
 111 11 : 1 11
 985 PVGSGPPKGDGASG 999

RT "Covalent structure of collagen: amino acid sequence of alpha
RT 1(III)-CB9 from type III collagen of human liver.";
RL *Biochemistry* 20:2621-2627(1981).

RN [11]

SEQUENCE OF 1176-1466 FROM N.A.
RX MEDLINE=85157600; PubMed=2279949;

RA Chu M.-L., Weil D., de Wet W.J., Bernard M.P., Sippola M., Ramirez F.;

RT "Isolation of cDNA and genomic clones encoding human pro-alpha 1
RT (III) collagen. Partial characterization of the 3' end region of the
RT gene.";
RL *J. Biol. Chem.* 260:4357-4363(1985).

RN [12]

SEQUENCE OF 1161-1200 FROM N.A.
RX MEDLINE=861187804; PubMed=3154462;

RA Miskulin M., Daigleish R., Klueve-Beckerman B., Rennard S.I.,
Tolstoshev P., Brantly M., Crystal R.G.;

RT "Human type III collagen gene expression is coordinately modulated
with the type I collagen genes during fibroblast growth.";
RL *Biochemistry* 25:1408-1413(1986).

RN [13]

SEQUENCE OF 1-170 FROM N.A.
RX TISSUE=placenta;
RX MEDLINE=88303360; PubMed=3405773;

RA Toman D., Ricca G., de Crombrughe B.;"
"Nucleotide sequence of cDNA coding for the amino-terminal region
of human prepro alpha 1(III) collagen.;"
RL *Nucleic Acids Res.* 16:7201-7201(1988).

RN [14]

SEQUENCE OF 1-176 FROM N.A.
RX MEDLINE=89378752; PubMed=2777083;

RA Benson-Chanda V., Su M.W., Weil D., Chu M.-L., Ramirez F.;
RT "Cloning and analysis of the 5' portion of the human type-III
procollagen gene (COL3A1).";
RL *Gene* 78:255-265(1989).

RN [15]

REVIEW ON VARIANTS:
RX MEDLINE=97255959; PubMed=9101290;

RA Kuivaniemi H., Tromp G., Prockop D.J.;
RT "Mutations in fibrillar collagens (types I, II, III, and XI), fibrill-
associated collagen (type IX), and network-forming collagen (type X),
cause a spectrum of diseases of bone, cartilage, and blood vessels.";
RL *Hum. Mutat.* 9:300-315(1997).

RN [16]

VARIANT AORTIC ANEURYSM ARG-303, AND VARIANT THR-668.
RX MEDLINE=93203988; PubMed=8114866;

RA Tromp G., Wu Y., Prockop D.J., Madhatteri S.L., Kleinert C.,
RA Earley J.J., Zhuang J., Noergaard O., Darling R.C., Abbott W.M.,
Cole C.W., Jaakkola P., Ryynanen M., Pearce W.H., Yao J.S.T.,
RA McJames K., Smulders S.N., Gatalica Z., Ferrell R.E., Jimenez S.A.,
RA Jackson C.E., Michels V.V., Kaye M., Kuivaniemi H.;"
RT "sequencing of cDNA from 50 unrelated patients reveals that mutations
in the triple-helical domain of type III procollagen are an
infrequent cause of aortic aneurysms.";
RL *J. Clin. Invest.* 91:2239-2245(1993).

RN [17]

VARIANT THR-698.
RX MEDLINE=91045136; PubMed=2235526;

RA Zafarullah K., Kleinert C., Tromp G., Kuivaniemi H., Kontusaari S.,
RA "G to A polymorphism in exon 31 of the COL3A1 gene.";
RL *Nucleic Acids Res.* 18:6180-6180(1990).

RN [18]

VARIANT AORTIC ANEURYSM ARG-786.
RX MEDLINE=91056145; PubMed=2243125;

RA Kontusaari S., Tromp G., Kuivaniemi H., Romanic A.M., Prockop D.J.;
RT "A mutation in the gene for type III procollagen (COL3A1) in a family
with aortic aneurysms.";
RL *J. Clin. Invest.* 86:1465-1473(1990).

RN [19]

VARIANT EDS-IV ARG-828.
RX MEDLINE=94016385; PubMed=8411057;

RA Richards A.J., Narcisi P., Lloyd J.C., Ferguson C., Pope F.M.,
Richards J., "The substitution of glycine 661 by arginine in type III collagen

RT produces mutant molecules with different thermal stabilities and
RT causes Ehlers-Danlos syndrome type IV.";
RL *J. Med. Genet.* 30:690-693(1993).

RN [20]

VARIANT EDS-IV SER-957. PubMed=2492273;

RX MEDLINE=89109135; PubMed=2279949;

RA Tromp G., Kuivaniemi H., Shikata H., Prockop D.J.;"
"A single base mutation that substitutes serine for glycine 790 of
the alpha 1(III) chain of type III procollagen exposes an arginine
and causes Ehlers-Danlos syndrome type IV.";
RL *J. Biol. Chem.* 264:1349-1352(1989).

RN [21]

VARIANT EDS-IV VAL-960.
RX MEDLINE=95208429; PubMed=7749417;

RA Tromp G., de Paepe A., Nyutnick L., Madhatteri S.L., Kuivaniemi H.,
Richards A.J., Ward P.N., Narcisi P., Nicholls A.C., Lloyd J.C.,
Pope F.M.;"
"Substitution of valine for glycine 793 in type III procollagen in
Ehlers-Danlos syndrome type IV.";
RL *Hum. Mutat.* 5:179-181(1995).

[22]

VARIANT EDS-IV GLU-1014.
RX MEDLINE=92316511; PubMed=1352273;

RA Tromp G., de Paepe A., Nyutnick L., Madhatteri S.L., Kuivaniemi H.,
Richards A.J., Ward P.N., Narcisi P., Nicholls A.C., Lloyd J.C.,
Pope F.M.;"
"A single base mutation in the gene for type III collagen (COL3A1)
converts glycine 847 to glutamic acid in a family with Ehlers-Danlos
syndrome type IV. An unaffected family member is mosaic for the
mutation.";
RL *Hum. Genet.* 89:414-418(1992).

RN [23]

VARIANT EDS-IV ASP-1050.
RX MEDLINE=90037070; PubMed=2808425;

RA Tromp G., Kuivaniemi H., Stolle C.A., Pope F.M., Prockop D.J.;
Richards A.J., Lloyd J.C., Ward P.N., de Paepe A., Narcisi P.,
Pope F.M.;"
"Single base mutation in the type III procollagen gene that converts
the codon for glycine 883 to aspartate in a mild variant of
Ehlers-Danlos syndrome type IV.";
RL *J. Biol. Chem.* 264:19113-19117(1989).

[24]

VARIANT EDS-IV VAL-1077. PubMed=1895316;

RX MEDLINE=91374480; PubMed=1895316;

RA Richards A.J., Lloyd J.C., Ward P.N., de Paepe A., Narcisi P.,
Pope F.M.;"
"Characterisation of a glycine to valine substitution at amino acid
position 910 of the triple helical region of type III collagen in a
patient with Ehlers-Danlos syndrome type IV.";
RL *J. Med. Genet.* 28:458-463(1991).

RN [25]

VARIANT EDS-IV GLU-1173.
RX MEDLINE=93022543; PubMed=1357232;

RA Johnson P.H., Richards A.J., Pope F.M., Hopkinson D.A.,
Richards A.J., Lloyd J.C., Ward P.N., de Paepe A., Narcisi P.,
Pope F.M.;"
"Query Match 58.7%; Score 54; DB 1; Length 1466;
Best Local Similarity 66.7%; Pred. No. 5.9; Matches 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0; Db 1141 PVGPSPGPKGDGTSG 1155

RESULT 25

CA1H_MOUSE STANDARD PRT: 1527 AA.

ID CA1H_MOUSE

AC P39061; 062002; 061437; Created

DT 01-FEB-1995 (Rel. 31, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DT 01-MAR-2002 (Rel. 41, Last annotation update)

DE Collagen alpha 1(XVIII) chain precursor [Contains: Endostatin],
GN Cor18A1.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TAXID=10090;

RN [1]

RP SEQUENCE FROM N.A. (SHORT ISOFORM).
 RC STRAIN=BALB/C; TISSUE=LIVER;
 RX MEDLINE=94245707; PubMed=8188673;
 RA Rehn M.V.; Hintikka E.; Piha-Jananiemi T.;
 RT "Primary structure of the alpha 1 chain of mouse type XVIII collagen,
 partial structure of the corresponding gene, and comparison of the
 alpha 1(XVIII) chain with its homologue, the alpha 1(XV) collagen
 chain.";
 J. Biol. Chem. 269:13929-13935(1994).

[2] SEQUENCE FROM N.A. (SHORT AND LONG ISOFORMS).
 RA Rehn M., Hintikka E., Piha-Jananiemi T.;
 RT "Characterization of the mouse gene for the alpha-1 chain of type
 XVIII collagen (COL8A1) reveals that the three variant N-terminal
 polypeptide forms are transcribed from two widely separated
 promoters.";
 RL Submitted (NOV-1993) to the EMBL/genBank/DDJB databases.
 RN [3]
 RP SEQUENCE OF 213-11402 FROM N.A. (SHORT ISOFORM).
 RX MEDLINE=94240112; PubMed=8183894;
 RA Rehn M.V.; Piha-Jananiemi T.;
 RT Alpha 1(XVIII), a collagen chain with frequent interruptions in the
 collagenous sequence, a distinct tissue distribution, and homology
 with type XV collagen.;
 RL Proc. Natl. Acad. Sci. U.S.A. 91:4234-4238(1994).
 RN [4]
 RP SEQUENCE OF 240-1527 FROM N.A.
 RX TISSUE=Liver;
 MEDLINE=94240111; PubMed=8183893;
 RA On S.P., Kamagata Y., Miragaki Y., Timmons S., Ooshima A., Olsen B.R.;
 RT "Isolation and sequencing of cDNAs for proteins with multiple domains
 of GLY-Xaa-Yaa repeats identify a distinct family of collagenous
 proteins";
 RL Proc. Natl. Acad. Sci. U.S.A. 91:4229-4233(1994).
 RN [5]
 RP CHARACTERIZATION OF ENDOSTATIN, AND PARTIAL SEQUENCE.
 RX MEDLINE=931160848; PubMed=9008168;
 RA O'Reilly M.S., Boehm T., Shing Y., Fukai N., Vassios G., Lane W.S.,
 Flynn E., Birkhead J.R., Olsen B.R., Folkman J.;
 RT "Endostatin: an endogenous inhibitor of angiogenesis and tumor
 growth.";
 RL Cell 88:277-285(1997).
 RN [6]
 RP X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS) OF ENDOSTATIN.
 RX MEDLINE=98169382; PubMed=9501087;
 RA Hohenester E., Saski T., Olsen B.R., Timpl R.;
 RT "Crystal structure of the angiogenesis inhibitor endostatin at 1.5-A
 resolution";
 RL EMBO J. 17:1656-1664(1998).
 CC -!- FUNCTION: ENDOSTATIN POTENTIALLY INHIBITS ENDOTHELIAL CELL
 PROLIFERATION AND ANGIOGENESIS. MAY INHIBIT ANGIOGENESIS BY
 BINDING TO THE REPARAN SULPHATE PROTEOGLYCANS INVOLVED IN GROWTH
 FACTOR SIGNALLING.
 CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A SHORT FORM AND A LONG FORM
 ("SHOWN HERE"); ARE PRODUCED BY ALTERNATIVE SPlicing.
 CC -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPETIDE REPEATING
 UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 CC -!- SIMILARITY: BELONGS TO THE FIBRIL-ASSOCIATED COLLAGENS WITH
 CC INTERRUPTED HELICES (FACT1) FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sb-sib.ch).
 CC -----
 DR EMBL: U34609; AAC52901.1; JOINED.
 DR EMBL: U34610; AAC52901.1; JOINED.
 DR EMBL: U34611; AAC52901.1; JOINED.
 DR EMBL: U34612; AAC52901.1; JOINED.
 DR EMBL: U34613; AAC52901.1; JOINED.
 DR EMBL: U34614; AAC52901.1; JOINED.
 DR EMBL: U34615; AAC52901.1; JOINED.
 DR EMBL: U34616; AAC52901.1; JOINED.
 DR EMBL: U34617; AAC52902.1; JOINED.
 DR EMBL: U34618; AAC52902.1; JOINED.
 DR EMBL: U34619; AAC52902.1; JOINED.
 DR EMBL: U34620; AAC52902.1; JOINED.
 DR EMBL: U34621; AAC52902.1; JOINED.
 DR EMBL: U34622; AAC52902.1; JOINED.
 DR EMBL: U34623; AAC52902.1; JOINED.
 DR EMBL: U34624; AAC52902.1; JOINED.
 DR EMBL: U34625; AAC52902.1; JOINED.
 DR EMBL: U34626; AAC52902.1; JOINED.
 DR EMBL: U34627; AAC52902.1; JOINED.
 DR EMBL: U34628; AAC52902.1; JOINED.
 DR EMBL: U34629; AAC52902.1; JOINED.
 DR EMBL: U34630; AAC52902.1; JOINED.
 DR EMBL: U34631; AAC52902.1; JOINED.
 DR EMBL: U34632; AAC52902.1; JOINED.
 DR EMBL: U34633; AAC52902.1; JOINED.
 DR EMBL: U34634; AAC52902.1; JOINED.
 DR EMBL: U34635; AAC52902.1; JOINED.
 DR EMBL: U34636; AAC52902.1; JOINED.
 DR EMBL: U34637; AAC52902.1; JOINED.
 DR EMBL: U34638; AAC52902.1; JOINED.
 DR EMBL: U34639; AAC52902.1; JOINED.
 DR EMBL: U34640; AAC52902.1; JOINED.
 DR EMBL: U34641; AAC52902.1; JOINED.
 DR EMBL: U34642; AAC52902.1; JOINED.
 DR EMBL: U34643; AAC52902.1; JOINED.
 DR EMBL: U34644; AAC52902.1; JOINED.
 DR EMBL: U34645; AAC52902.1; JOINED.
 DR EMBL: U34646; AAC52902.1; JOINED.
 DR EMBL: U34647; AAC52902.1; JOINED.
 DR EMBL: U34648; AAC52902.1; JOINED.
 DR EMBL: U34649; AAC52902.1; JOINED.
 DR EMBL: U34650; AAC52902.1; JOINED.
 DR EMBL: U34651; AAC52902.1; JOINED.
 DR EMBL: U34652; AAC52902.1; JOINED.
 DR EMBL: U34653; AAC52902.1; JOINED.
 DR EMBL: U34654; AAC52902.1; JOINED.
 DR EMBL: U34655; AAC52902.1; JOINED.
 DR EMBL: U34656; AAC52902.1; JOINED.
 DR EMBL: U34657; AAC52902.1; JOINED.
 DR EMBL: U34658; AAC52902.1; JOINED.
 DR EMBL: U34659; AAC52902.1; JOINED.
 DR EMBL: U34660; AAC52902.1; JOINED.
 DR EMBL: U34661; AAC52902.1; JOINED.
 DR EMBL: U34662; AAC52902.1; JOINED.
 DR EMBL: U34663; AAC52902.1; JOINED.
 DR EMBL: U34664; AAC52902.1; JOINED.
 DR EMBL: U34665; AAC52902.1; JOINED.
 DR EMBL: U34666; AAC52902.1; JOINED.
 DR EMBL: U34667; AAC52902.1; JOINED.
 DR EMBL: U34668; AAC52902.1; JOINED.
 DR EMBL: U34669; AAC52902.1; JOINED.
 DR EMBL: U34670; AAC52902.1; JOINED.
 DR EMBL: U34671; AAC52902.1; JOINED.
 DR EMBL: U34672; AAC52902.1; JOINED.
 DR EMBL: U34673; AAC52902.1; JOINED.
 DR EMBL: U34674; AAC52902.1; JOINED.
 DR EMBL: U34675; AAC52902.1; JOINED.
 DR EMBL: U34676; AAC52902.1; JOINED.
 DR EMBL: U34677; AAC52902.1; JOINED.
 DR EMBL: U34678; AAC52902.1; JOINED.
 DR EMBL: U34679; AAC52902.1; JOINED.
 DR EMBL: U34680; AAC52902.1; JOINED.
 DR EMBL: U34681; AAC52902.1; JOINED.
 DR EMBL: U34682; AAC52902.1; JOINED.
 DR EMBL: U34683; AAC52902.1; JOINED.
 DR EMBL: U34684; AAC52902.1; JOINED.
 DR EMBL: U34685; AAC52902.1; JOINED.
 DR EMBL: U34686; AAC52902.1; JOINED.
 DR EMBL: U34687; AAC52902.1; JOINED.
 DR EMBL: U34688; AAC52902.1; JOINED.
 DR EMBL: U34689; AAC52902.1; JOINED.
 DR EMBL: U34690; AAC52902.1; JOINED.
 DR EMBL: U34691; AAC52902.1; JOINED.
 DR EMBL: U34692; AAC52902.1; JOINED.
 DR EMBL: U34693; AAC52902.1; JOINED.
 DR EMBL: U34694; AAC52902.1; JOINED.
 DR EMBL: U34695; AAC52902.1; JOINED.
 DR EMBL: U34696; AAC52902.1; JOINED.
 DR EMBL: U34697; AAC52902.1; JOINED.
 DR EMBL: U34698; AAC52902.1; JOINED.
 DR EMBL: U34699; AAC52902.1; JOINED.
 DR EMBL: U34700; AAC52902.1; JOINED.
 DR EMBL: U34701; AAC52902.1; JOINED.
 DR EMBL: U34702; AAC52902.1; JOINED.
 DR EMBL: U34703; AAC52902.1; JOINED.
 DR EMBL: U34704; AAC52902.1; JOINED.
 DR EMBL: U34705; AAC52902.1; JOINED.
 DR EMBL: U34706; AAC52902.1; JOINED.
 DR EMBL: U34707; AAC52902.1; JOINED.
 DR EMBL: U34708; AAC52902.1; JOINED.
 DR EMBL: U34709; AAC52902.1; JOINED.
 DR EMBL: U34710; AAC52902.1; JOINED.
 DR EMBL: U34711; AAC52902.1; JOINED.
 DR EMBL: U34712; AAC52902.1; JOINED.
 DR EMBL: U34713; AAC52902.1; JOINED.
 DR EMBL: U34714; AAC52902.1; JOINED.
 DR EMBL: U34715; AAC52902.1; JOINED.
 DR EMBL: U34716; AAC52902.1; JOINED.
 DR EMBL: U34717; AAC52902.1; JOINED.
 DR EMBL: U34718; AAC52902.1; JOINED.
 DR EMBL: U34719; AAC52902.1; JOINED.
 DR EMBL: U34720; AAC52902.1; JOINED.
 DR EMBL: U34721; AAC52902.1; JOINED.
 DR EMBL: U34722; AAC52902.1; JOINED.
 DR EMBL: U34723; AAC52902.1; JOINED.
 DR EMBL: U34724; AAC52902.1; JOINED.
 DR EMBL: U34725; AAC52902.1; JOINED.
 DR EMBL: U34726; AAC52902.1; JOINED.
 DR EMBL: U34727; AAC52902.1; JOINED.
 DR EMBL: U34728; AAC52902.1; JOINED.
 DR EMBL: U34729; AAC52902.1; JOINED.
 DR EMBL: U34730; AAC52902.1; JOINED.
 DR EMBL: U34731; AAC52902.1; JOINED.
 DR EMBL: U34732; AAC52902.1; JOINED.
 DR EMBL: U34733; AAC52902.1; JOINED.
 DR EMBL: U34734; AAC52902.1; JOINED.
 DR EMBL: U34735; AAC52902.1; JOINED.
 DR EMBL: U34736; AAC52902.1; JOINED.
 DR EMBL: U34737; AAC52902.1; JOINED.
 DR EMBL: U34738; AAC52902.1; JOINED.
 DR EMBL: U34739; AAC52902.1; JOINED.
 DR EMBL: U34740; AAC52902.1; JOINED.
 DR EMBL: U34741; AAC52902.1; JOINED.
 DR EMBL: U34742; AAC52902.1; JOINED.
 DR EMBL: U34743; AAC52902.1; JOINED.
 DR EMBL: U34744; AAC52902.1; JOINED.
 DR EMBL: U34745; AAC52902.1; JOINED.
 DR EMBL: U34746; AAC52902.1; JOINED.
 DR EMBL: U34747; AAC52902.1; JOINED.
 DR EMBL: U34748; AAC52902.1; JOINED.
 DR EMBL: U34749; AAC52902.1; JOINED.
 DR EMBL: U34750; AAC52902.1; JOINED.
 DR EMBL: U34751; AAC52902.1; JOINED.
 DR EMBL: U34752; AAC52902.1; JOINED.
 DR EMBL: U34753; AAC52902.1; JOINED.
 DR EMBL: U34754; AAC52902.1; JOINED.
 DR EMBL: U34755; AAC52902.1; JOINED.
 DR EMBL: U34756; AAC52902.1; JOINED.
 DR EMBL: U34757; AAC52902.1; JOINED.
 DR EMBL: U34758; AAC52902.1; JOINED.
 DR EMBL: U34759; AAC52902.1; JOINED.
 DR EMBL: U34760; AAC52902.1; JOINED.
 DR EMBL: U34761; AAC52902.1; JOINED.
 DR EMBL: U34762; AAC52902.1; JOINED.
 DR EMBL: U34763; AAC52902.1; JOINED.
 DR EMBL: U34764; AAC52902.1; JOINED.
 DR EMBL: U34765; AAC52902.1; JOINED.
 DR EMBL: U34766; AAC52902.1; JOINED.
 DR EMBL: U34767; AAC52902.1; JOINED.
 DR EMBL: U34768; AAC52902.1; JOINED.
 DR EMBL: U34769; AAC52902.1; JOINED.
 DR EMBL: U34770; AAC52902.1; JOINED.
 DR EMBL: U34771; AAC52902.1; JOINED.
 DR EMBL: U34772; AAC52902.1; JOINED.
 DR EMBL: U34773; AAC52902.1; JOINED.
 DR EMBL: U34774; AAC52902.1; JOINED.
 DR EMBL: U34775; AAC52902.1; JOINED.
 DR EMBL: U34776; AAC52902.1; JOINED.
 DR EMBL: U34777; AAC52902.1; JOINED.
 DR EMBL: U34778; AAC52902.1; JOINED.
 DR EMBL: U34779; AAC52902.1; JOINED.
 DR EMBL: U34780; AAC52902.1; JOINED.
 DR EMBL: U34781; AAC52902.1; JOINED.
 DR EMBL: U34782; AAC52902.1; JOINED.
 DR EMBL: U34783; AAC52902.1; JOINED.
 DR EMBL: U34784; AAC52902.1; JOINED.
 DR EMBL: U34785; AAC52902.1; JOINED.
 DR EMBL: U34786; AAC52902.1; JOINED.
 DR EMBL: U34787; AAC52902.1; JOINED.
 DR EMBL: U34788; AAC52902.1; JOINED.
 DR EMBL: U34789; AAC52902.1; JOINED.
 DR EMBL: U34790; AAC52902.1; JOINED.
 DR EMBL: U34791; AAC52902.1; JOINED.
 DR EMBL: U34792; AAC52902.1; JOINED.
 DR EMBL: U34793; AAC52902.1; JOINED.
 DR EMBL: U34794; AAC52902.1; JOINED.
 DR EMBL: U34795; AAC52902.1; JOINED.
 DR EMBL: U34796; AAC52902.1; JOINED.
 DR EMBL: U34797; AAC52902.1; JOINED.
 DR EMBL: U34798; AAC52902.1; JOINED.
 DR EMBL: U34799; AAC52902.1; JOINED.
 DR EMBL: U34800; AAC52902.1; JOINED.
 DR EMBL: U34801; AAC52902.1; JOINED.
 DR EMBL: U34802; AAC52902.1; JOINED.
 DR EMBL: U34803; AAC52902.1; JOINED.
 DR EMBL: U34804; AAC52902.1; JOINED.
 DR EMBL: U34805; AAC52902.1; JOINED.
 DR EMBL: U34806; AAC52902.1; JOINED.
 DR EMBL: U34807; AAC52902.1; JOINED.
 DR EMBL: U34808; AAC52902.1; JOINED.
 DR EMBL: U34809; AAC52902.1; JOINED.
 DR EMBL: U34810; AAC52902.1; JOINED.
 DR EMBL: U34811; AAC52902.1; JOINED.
 DR EMBL: U34812; AAC52902.1; JOINED.
 DR EMBL: U34813; AAC52902.1; JOINED.
 DR EMBL: U34814; AAC52902.1; JOINED.
 DR EMBL: U34815; AAC52902.1; JOINED.
 DR EMBL: U34816; AAC52902.1; JOINED.
 DR EMBL: U34817; AAC52902.1; JOINED.
 DR EMBL: U34818; AAC52902.1; JOINED.
 DR EMBL: U34819; AAC52902.1; JOINED.
 DR EMBL: U34820; AAC52902.1; JOINED.
 DR EMBL: U34821; AAC52902.1; JOINED.
 DR EMBL: U34822; AAC52902.1; JOINED.
 DR EMBL: U34823; AAC52902.1; JOINED.
 DR EMBL: U34824; AAC52902.1; JOINED.
 DR EMBL: U34825; AAC52902.1; JOINED.
 DR EMBL: U34826; AAC52902.1; JOINED.
 DR EMBL: U34827; AAC52902.1; JOINED.
 DR EMBL: U34828; AAC52902.1; JOINED.
 DR EMBL: U34829; AAC52902.1; JOINED.
 DR EMBL: U34830; AAC52902.1; JOINED.
 DR EMBL: U34831; AAC52902.1; JOINED.
 DR EMBL: U34832; AAC52902.1; JOINED.
 DR EMBL: U34833; AAC52902.1; JOINED.
 DR EMBL: U34834; AAC52902.1; JOINED.
 DR EMBL: U34835; AAC52902.1; JOINED.
 DR EMBL: U34836; AAC52902.1; JOINED.
 DR EMBL: U34837; AAC52902.1; JOINED.
 DR EMBL: U34838; AAC52902.1; JOINED.
 DR EMBL: U34839; AAC52902.1; JOINED.
 DR EMBL: U34840; AAC52902.1; JOINED.
 DR EMBL: U34841; AAC52902.1; JOINED.
 DR EMBL: U34842; AAC52902.1; JOINED.
 DR EMBL: U34843; AAC52902.1; JOINED.
 DR EMBL: U34844; AAC52902.1; JOINED.
 DR EMBL: U34845; AAC52902.1; JOINED.
 DR EMBL: U34846; AAC52902.1; JOINED.
 DR EMBL: U34847; AAC52902.1; JOINED.
 DR EMBL: U34848; AAC52902.1; JOINED.
 DR EMBL: U34849; AAC52902.1; JOINED.
 DR EMBL: U34850; AAC52902.1; JOINED.
 DR EMBL: U34851; AAC52902.1; JOINED.
 DR EMBL: U34852; AAC52902.1; JOINED.
 DR EMBL: U34853; AAC52902.1; JOINED.
 DR EMBL: U34854; AAC52902.1; JOINED.
 DR EMBL: U34855; AAC52902.1; JOINED.
 DR EMBL: U34856; AAC52902.1; JOINED.
 DR EMBL: U34857; AAC52902.1; JOINED.
 DR EMBL: U34858; AAC52902.1; JOINED.
 DR EMBL: U34859; AAC52902.1; JOINED.
 DR EMBL: U34860; AAC52902.1; JOINED.
 DR EMBL: U34861; AAC52902.1; JOINED.
 DR EMBL: U34862; AAC52902.1; JOINED.
 DR EMBL: U34863; AAC52902.1; JOINED.
 DR EMBL: U34864; AAC52902.1; JOINED.
 DR EMBL: U34865; AAC52902.1; JOINED.
 DR EMBL: U34866; AAC52902.1; JOINED.
 DR EMBL: U34867; AAC52902.1; JOINED.
 DR EMBL: U34868; AAC52902.1; JOINED.
 DR EMBL: U34869; AAC52902.1; JOINED.
 DR EMBL: U34870; AAC52902.1; JOINED.
 DR EMBL: U34871; AAC52902.1; JOINED.
 DR EMBL: U34872; AAC52902.1; JOINED.
 DR EMBL: U34873; AAC52902.1; JOINED.
 DR EMBL: U34874; AAC52902.1; JOINED.
 DR EMBL: U34875; AAC52902.1; JOINED.
 DR EMBL: U34876; AAC52902.1; JOINED.
 DR EMBL: U34877; AAC52902.1; JOINED.
 DR EMBL: U34878; AAC52902.1; JOINED.
 DR EMBL: U34879; AAC52902.1; JOINED.
 DR EMBL: U34880; AAC52902.1; JOINED.
 DR EMBL: U34881; AAC52902.1; JOINED.
 DR EMBL: U34882; AAC52902.1; JOINED.
 DR EMBL: U34883; AAC52902.1; JOINED.
 DR EMBL: U34884; AAC52902.1; JOINED.
 DR EMBL: U34885; AAC52902.1; JOINED.
 DR EMBL: U34886; AAC52902.1; JOINED.
 DR EMBL: U34887; AAC52902.1; JOINED.
 DR EMBL: U34888; AAC52902.1; JOINED.
 DR EMBL: U34889; AAC52902.1; JOINED.
 DR EMBL: U34890; AAC52902.1; JOINED.
 DR EMBL: U34891; AAC52902.1; JOINED.
 DR EMBL: U34892; AAC52902.1; JOINED.
 DR EMBL: U34893; AAC52902.1; JOINED.
 DR EMBL: U34894; AAC52902.1; JOINED.
 DR EMBL: U34895; AAC52902.1; JOINED.
 DR EMBL: U34896; AAC52902.1; JOINED.
 DR EMBL: U34897; AAC52902.1; JOINED.
 DR EMBL: U34898; AAC52902.1; JOINED.
 DR EMBL: U34899; AAC52902.1; JOINED.
 DR EMBL: U34900; AAC52902.1; JOINED.
 DR EMBL: U34901; AAC52902.1; JOINED.
 DR EMBL: U34902; AAC52902.1; JOINED.
 DR EMBL: U34903; AAC52902.1; JOINED.
 DR EMBL: U34904; AAC52902.1; JOINED.
 DR EMBL: U34905; AAC52902.1; JOINED.
 DR EMBL: U34906; AAC52902.1; JOINED.
 DR EMBL: U34907; AAC52902.1; JOINED.
 DR EMBL: U34908; AAC52902.1; JOINED.
 DR EMBL: U34909; AAC52902.1; JOINED.
 DR EMBL: U34910; AAC52902.1; JOINED.
 DR EMBL: U34911; AAC52902.1; JOINED.
 DR EMBL: U34912; AAC52902.1; JOINED.
 DR EMBL: U34913; AAC52902.1; JOINED.
 DR EMBL: U34914; AAC52902.1; JOINED.
 DR EMBL: U34915; AAC52902.1; JOINED.
 DR EMBL: U34916; AAC52902.1; JOINED.
 DR EMBL: U34917; AAC52902.1; JOINED.
 DR EMBL: U34918; AAC52902.1; JOINED.
 DR EMBL: U34919; AAC52902.1; JOINED.
 DR EMBL: U34920; AAC52902.1; JOINED.
 DR EMBL: U34921; AAC52902.1; JOINED.
 DR EMBL: U34922; AAC52902.1; JOINED.
 DR EMBL: U34923; AAC52902.1; JOINED.
 DR EMBL: U34924; AAC52902.1; JOINED.
 DR EMBL: U34925; AAC52902.1; JOINED.
 DR EMBL: U34926; AAC52902.1; JOINED.
 DR EMBL: U34927; AAC52902.1; JOINED.
 DR EMBL: U34928; AAC52902.1; JOINED.
 DR EMBL: U34929; AAC52902.1; JOINED.
 DR EMBL: U34930; AAC52902.1; JOINED.
 DR EMBL: U34931; AAC52902.1; JOINED.
 DR EMBL: U34932; AAC52902.1; JOINED.
 DR EMBL: U34933; AAC52902.1; JOINED.
 DR EMBL: U34934; AAC52902.1; JOINED.
 DR EMBL: U34935; AAC52902.1; JOINED.
 DR EMBL: U34936; AAC52902.1; JOINED.
 DR EMBL: U34937; AAC52902.1; JOINED.
 DR EMBL: U34938; AAC52902.1; JOINED.
 DR EMBL: U34939; AAC52902.1; JOINED.
 DR EMBL: U34940; AAC52902.1; JOINED.
 DR EMBL: U34941; AAC52902.1; JOINED.
 DR EMBL: U34942; AAC52902.1; JOINED.
 DR EMBL: U34943; AAC52902.1; JOINED.
 DR EMBL: U34944; AAC52902.1; JOINED.
 DR EMBL: U34945; AAC52902.1; JOINED.
 DR EMBL: U34946; AAC52902.1; JOINED.
 DR EMBL: U34947; AAC52902.1; JOINED.
 DR EMBL: U34948; AAC52902.1; JOINED.
 DR EMBL: U34949; AAC52902.1; JOINED.
 DR EMBL: U34950; AAC52902.1; JOINED.
 DR EMBL: U34951; AAC52902.1; JOINED.
 DR EMBL: U34952; AAC52902.1; JOINED.
 DR EMBL: U34953; AAC52902.1; JOINED.
 DR EMBL: U34954; AAC52902.1; JOINED.
 DR EMBL: U34955; AAC52902.1; JOINED.
 DR EMBL: U34956; AAC52902.1; JOINED.
 DR EMBL: U34957; AAC52902.1; JOINED.
 DR EMBL: U34958; AAC52902.1; JOINED.
 DR EMBL: U34959; AAC52902.1; JOINED.
 DR EMBL: U34960; AAC52902.1; JOINED.
 DR EMBL: U34961; AAC52902.1; JOINED.
 DR EMBL: U34962; AAC52902.1; JOINED.
 DR EMBL: U34963; AAC52902.1; JOINED.
 DR EMBL: U34964; AAC52902.1; JOINED.
 DR EMBL: U34965; AAC52902.1; JOINED.
 DR EMBL: U34966; AAC52902.1; JOINED.
 DR EMBL: U34967; AAC52902.1; JOINED.
 DR EMBL: U34968; AAC52902.1; JOINED.
 DR EMBL: U34969; AAC52902.1; JOINED.
 DR EMBL: U34970; AAC52902.1; JOINED.
 DR EMBL: U34971; AAC52902.1; JOINED.
 DR EMBL: U34972; AAC52902.1; JOINED.
 DR EMBL: U34973; AAC52902.1; JOINED.
 DR EMBL: U34974; AAC52902.1; JOINED.
 DR EMBL: U34975; AAC52902.1; JOINED.
 DR EMBL: U34976; AAC52902.

PT MOD_RES 774 774 HYDROXYLATION.
 PT MOD_RES 780 780 HYDROXYLATION.
 PT MOD_RES 789 789 HYDROXYLATION.
 PT MOD_RES 795 795 HYDROXYLATION.
 PT MOD_RES 804 804 HYDROXYLATION.
 PT MOD_RES 807 807 HYDROXYLATION.
 PT MOD_RES 810 810 HYDROXYLATION.
 PT MOD_RES 816 816 HYDROXYLATION.
 PT MOD_RES 819 819 HYDROXYLATION.
 PT MOD_RES 834 834 HYDROXYLATION.
 PT MOD_RES 846 846 HYDROXYLATION.
 PT MOD_RES 861 861 HYDROXYLATION.
 PT MOD_RES 864 864 HYDROXYLATION.
 PT MOD_RES 870 870 HYDROXYLATION.
 PT MOD_RES 873 873 HYDROXYLATION.
 PT MOD_RES 906 906 HYDROXYLATION.
 PT MOD_RES 930 930 HYDROXYLATION.
 PT MOD_RES 945 945 HYDROXYLATION.
 PT MOD_RES 1017 1017 HYDROXYLATION.
 PT MOD_RES 1020 1020 HYDROXYLATION.
 PT MOD_RES 1023 1023 HYDROXYLATION.
 PT MOD_RES 1029 1029 HYDROXYLATION.
 PT MOD_RES 1221 1221 HYDROXYLATION.
 PT MOD_RES 1224 1224 HYDROXYLATION.
 PT MOD_RES 1467 1467 HYDROXYLATION.
 PT MOD_RES 1470 1470 HYDROXYLATION.
 PT VARIANT 1639 1639 C -> S (IN EDSL).
 />FTid:VAR_001808.
 PT CONFLICT 641 641 E -> G (IN REF. 2).
 PT CONFLICT 650 650 P -> L (IN REF. 2).
 PT CONFLICT 663 663 R -> E (IN REF. 2).
 PT CONFLICT 668 668 E -> O (IN REF. 2).
 PT CONFLICT 677 677 K -> Q (IN REF. 2).
 PT CONFLICT 684 684 L -> P (IN REF. 2).
 PPGPQVTT -> VTGGERGAP (IN REF. 2).
 PT CONFLICT 692 692 P -> Q (IN REF. 2).
 PT CONFLICT 727 727 P -> L (IN REF. 2).
 PT CONFLICT 741 741 P -> O (IN REF. 2).
 PT CONFLICT 747 747 L -> P (IN REF. 2).
 PT CONFLICT 753 753 P -> A (IN REF. 2).
 PT CONFLICT 759 759 D -> N (IN REF. 2).
 GO -> OK (IN REF. 2).
 GGPNGDP -> IGGPPPR (IN REF. 3).
 PT CONFLICT 849 855 Query Match 57.6%; Score 53; DB 1; Length 307;
 SEQUENCE 1838 AA; 183616 MW; 7D58239C0D77BDAE CRC64;
 Best Local Similarity 58.7%; Score 54; DB 1; Length 1838;
 Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
 SQ [1]
 QY 1 PAGPVGPNKGDKGVG 15
 ||| ||| | :| |
 Db 1145 PAGPVGPPGEDGKG 1159
 RESULT 27
 CC36_CAEEL STANDARD; PRT; 307 AA.
 ID CC06_CAEEL STANDARD; PRT; 329 AA.
 AC P34803;
 DT 01-FEB-1994 (Rel. 28 Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Cuticle collagen 36.
 COL-36 OR C27H5.5.
 OS Caenorhabditis elegans.
 Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 Rhabditidae; peloderaiae; Caenorhabditis.
 NCBI_TaxID=6239;

RN [1] SEQUENCE FROM N.A.
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL_N2;
 RX MEDLINE=9413128; PubMed=8299960;
 RA Levy A.D., Kramer J.M.;
 RT "Identification, sequence and expression patterns of the
 Caenorhabditis elegans col-36 and col-40 collagen-encoding genes.",
 RL Gene 137:281-285(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL_N2;
 RA Pauley A.;
 RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
 CC !- FUNCTION: NEMATODE CUTICLES ARE COMPOSED LARGELY OF COLLAGEN-LIKE
 PROTEINS. THE CUTICLE FUNCTIONS BOTH AS AN EXOSKELETON AND AS A
 BARRIER TO PROTECT THE WORM FROM ITS ENVIRONMENT.
 CC -! SUBUNIT: COLLAGEN POLYPEPTIDE CHAINS ARE COMPLEXED WITHIN THE
 CUTICLE BY DISULFIDE BONDS AND OTHER TYPES OF COVALENT CROSS-
 LINKS.
 CC -! SIMILARITY: TO OTHER COLLAGENS. STRONG, TO OTHER CUTICLE
 CC -! COLLAGENS.

This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

DR EMBL; L15418; AAA17445.1; .
 DR EMBL; U14635; AAK6492.1; .
 DR WormBase; C27H5.5; CEG6893.
 DR InterPro; IPR002486; Col_cuticle_N.
 DR Pfam; PF01391; Collagen; 3.
 DR Pfam; PF01484; Col_cuticle_N; 1.
 KW Cuticle; Connective tissue; Repeat; Multigene family; Collagen.
 FT DOMAIN 89 105 TRIPLE-HELICAL REGION.
 FT DOMAIN 118 150 TRIPLE-HELICAL REGION.
 FT DOMAIN 157 187 TRIPLE-HELICAL REGION.
 FT DOMAIN 194 226 TRIPLE-HELICAL REGION.
 FT DOMAIN 231 257 TRIPLE-HELICAL REGION.
 FT DOMAIN 260 295 TRIPLE-HELICAL REGION.
 SQ SEQUENCE 307 AA; 30126 MW; 9C7206CC18660F0B CRC64;

Query Match 57.6%; Score 53; DB 1; Length 307;
 Best Local Similarity 56.7%; Pred. No. 1.9; 5; Indels 0; Gaps 0;
 Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
 SQ [1]
 QY 1 PAGPVGPNKGDKGVG 15
 ||| ||| | :| |
 Db 125 PAGDVGPNGDGHG 139

RESULT 28
 CC06_CAEEL STANDARD; PRT; 329 AA.
 ID CC06_CAEEL STANDARD; PRT; 329 AA.
 AC P18831;
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 01-NOV-1990 (Rel. 16, Last sequence update)
 DE Cuticle collagen 6.
 GN COL-5.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderaiae; Caenorhabditis.
 OC NCBI_TaxID=6239;
 RN [1] SEQUENCE FROM N.A.
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL_N2;
 RX MEDLINE=89326131; PubMed=2753356;
 RA Cox G.N., Fields C., Kramer J.M., Rosenzweig B., Hirsh D.;

RT *sequence comparisons of developmentally regulated collagen genes of
 RT Caenorhabditis elegans.";
 RL Gene 76:331-344(1989).
 CC -!- FUNCTION: NEMATODE CUTICLES ARE COMPOSED LARGELY OF COLLAGEN-LIKE
 PROTEINS. THE CUTICLE FUNCTIONS BOTH AS AN EXOSKELETON AND AS A
 CC BARRIER TO PROTECT THE WORM FROM ITS ENVIRONMENT.
 CC -!- SUBUNIT: COLLAGEN POLYPEPTIDE CHAINS ARE COMPLEXED WITHIN THE
 CUTICLE BY DISULFIDE BONDS AND OTHER TYPES OF COVALENT CROSS-
 CC LINKS.
 CC -!- SIMILARITY: TO OTHER COLLAGENS. STRONG TO OTHER CUTICLE
 CC COLLAGENS. COL-2 AND COL-6 BELONGS TO THE SAME GROUP OF COLLAGEN.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; M25477; AA227991.1; -.
 DR InterPro; IPR002486; Col-cuticle_N.
 DR InterPro; IPR00087; Collagen.
 DR Pfam; PF01391; Collagen; 2.
 DR Pfam; PF01448; Col-cuticle_N; 1.
 KW Curcicle; Connective tissue; Repeat; Multigene family; Collagen.
 FT DOMAIN 95 115 GY-RICH
 FT DOMAIN 142 171 TRIPLE-HELICAL REGION.
 FT DOMAIN 149 212 TRIPLE-HELICAL REGION.
 FT DOMAIN 216 248 TRIPLE-HELICAL REGION.
 FT DOMAIN 253 279 TRIPLE-HELICAL REGION.
 FT DOMAIN 282 320 TRIPLE-HELICAL REGION.
 SQ SEQUENCE 329 AA; 32561 MW; 52C295EF77E76418 CRC64;

Query Match 57.6%; Score 53; DB 1; Length 329;
 Best Local Similarity 66.7%; Pred. No. 2;
 Matches 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 QY 1 PAGPWGNGKDGKV 15
 Db 146 PROPEGPQGNDGKD 160

RESULT 29

ID	PSPD_MOUSE	STANDARD:	PRT:	374 AA.
PSPD_0404;				01-OCT-1996 (Rel. 34, Created)
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	15-JUL-1999 (Rel. 38, Last annotation update)			
DE	Pulmonary surfactant-associated protein D precursor (SP-D) (PSP-D).			
GN	SFPPD OR SFPP4.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=B6/CBA; TISSUE=Lung;			
RX	MEDLINE=96034460; PubMed=749852;			
RA	Motwani M., White R.A., Guo N., Dowler L.L., Tauber A.I., Sastry K.N.;			
RT	"Mouse surfactant protein-D, cDNA cloning, characterization, and gene localization to chromosome 14."			
RL	J. Immunol. 155:5671-5677(1995).			
CC	-!- FUNCTION: CONTRIBUTES TO THE LUNG'S DEFENSE AGAINST INHALED MICROORGANISMS. BINDS STRONGLY MALTOSE RESIDUES AND TO A LESSER EXTENT OTHER ALPHA-GLUCOSYL MOIETIES. IT COULD PARTICIPATE IN THE EXTRACELLULAR REORGANIZATION OR TURNOVER OF PULMONARY SURFACTANT.			
CC	-!- SUBUNIT: OLIGOMERIC COMPLEX OF 4 SET OF HOMOTRIMERS.			
CC	-!- SUBCELLULAR LOCATION: Extracellular.			
CC	-!- MISCELLANEOUS: PULMONARY SURFACTANT CONSISTS OF 90% LIPID AND 10%			
PROTEIN. THERE ARE 4 SURFACTANT ASSOCIATED PROTEIN: 2 COLLAGENOUS,				

RESULT 30

ID	PSPD_RAT	STANDARD:	PRT:	374 AA.
AC	P35248;			
DT	01-FEB-1994 (Rel. 28, Created)			
DT	01-FEB-1994 (Rel. 28, Last sequence update)			
DT	15-JUL-1999 (Rel. 38, Last annotation update)			
DE	Pulmonary surfactant-associated protein D precursor (SP-D) (PSP-D)			
DE	(CP4).			
GN	SFPPD OR SFPP4.			
OS	Rattus norvegicus (Rat).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
OX	NCBI_TaxID=10116;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 20-33.			
RC	TISSUE=Lung;			
RX	MEDLINE=9212913; PubMed=1370483;			
RA	Shimizu H., Fisher J.H., Popst P., Benson B., Lau K., Mason R.J., Voelker D.R.;			
RA	"Primary structure of rat pulmonary surfactant protein D. cDNA and RT deduced amino acid sequence";			
RL	J. Biol. Chem. 267:1855-1857(1992).			
RN	[2]			
RP	SEQUENCE OF 73-95 AND 153-180.			
RC	TISSUE=Lung;			
RX	MEDLINE=9001186; PubMed=2675969;			
RA	Person A., Chang D., Rust K., Maxley M., Longmore W., Crouch E.;			

RT Purification and biochemical characterization of CP4 (SP-D), a
 RT collagenous surfactant-associated protein.";
 RL Biochemistry 28:6361-6367 (1999).
 CC -!- FUNCTION: CONTRIBUTES TO THE LUNG'S DEFENSE AGAINST INHALED
 CC MICROORGANISMS. BINDS STRONGLY MALTOSE RESIDUES AND TO A LESSER
 CC EXTENT OTHER ALPHA-GLUCOSYL MOIETIES. IT COULD PARTICIPATE IN THE
 CC EXTRACELLULAR REORGANIZATION OR TURNOVER OF PULMONARY SURFACTANT.
 CC -!- SUBUNIT: OLIGOMERIC COMPLEX OF 4 SET OF HOMOTRIMERS.
 CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
 CC -!- MISCELLANEOUS: PULMONARY SURFACTANT CONSISTS OF 90% LIPID AND 10%
 CC PROTEIN. THERE ARE 4 SURFACTANT ASSOCIATED PROTEIN: 2 COLLAGENOUS,
 CC CARBOHYDRATE-BINDING GLYcoproteins (SP-A AND SP-D) AND 2 SMALL
 CC HYDROPHOBIC PROTEINS (SP-B AND SP-C).
 CC -!- SIMILARITY: CONTAINS A COLLAGENOUS DOMAIN.
 CC -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; M81231; AAA42170.1; -.
 DR PIR; A42046; A42046.
 DR HSSP; P3247; 1B08.
 DR InterPro; IPR000087; Collagen.
 DR InterPro; IPR001304; lectin_c.
 DR Pfam; PF01391; Collagen; 3.
 DR Pfam; PF00059; lectin_c; 1.
 DR SMART; SM0034; CLSGC; 1.
 DR PROSITE; PS00615; C-TYPELECTIN_1; 1.
 DR PROSITE; PS5001; C-TYPELECTIN_2; 1.
 KW Glycoprotein; Calcium; Surface film; Gaseous exchange; Hydroxylation;
 KW Signal; Lectin; Collagen; Repeat; Coiled coil.
 FT SIGNAL 1 19
 FT CHAIN 20 374 PULMONARY SURFACTANT-ASSOCIATED PROTEIN
 FT
 FT DOMAIN 45 221 D. COLLAGEN-LIKE.
 FT DOMAIN 222 253 COILED COIL (POTENTIAL).
 FT DOMAIN 278 374 C-TYPE LECTIN (SHORT FORM).
 FT DISUFLID 280 372 BY SIMILARITY.
 FT DISUFLID 350 364 N-LINKED (GLCNAC. . .).
 FT CARBOHYD 89 89 HYDROXYLATION.
 FT MOD_RES 77 77 HYDROXYLATION.
 FT MOD_RES 86 86 HYDROXYLATION.
 FT MOD_RES 95 95 HYDROXYLATION.
 FT MOD_RES 98 98 HYDROXYLATION.
 FT MOD_RES 170 170 HYDROXYLATION.
 FT MOD_RES 176 176 HYDROXYLATION.
 FT CONFLICT 89 89 N -> E (IN REF. 2).
 FT CONFLICT 164 164 K -> C (IN REF. 2).
 SQ SEQUENCE 374 AA; 37561 MW; DB2BB5E3299DBA3C CRC64;
 Query Match 57.6%; Score 53; DB 1; Length 374;
 Best Local Similarity 60.0%; Pred. No. 2.3; Mismatches 4; Indels 0; Gaps 0;
 Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 1 PAGWGPNKGDKGVG 15

Db 115 PAGEGPSGKQGNIG 129

Search completed: November 1, 2002, 12:52:54
 Job time : 8 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 1, 2002, 12:19:21 ; Search time 21.5 Seconds
(without alignments)
120.694 Million cell updates/sec

Title: Perfect score: US-09-529-691A-3
Sequence: 1 PAGPWPSPNGKDKKVG 15

Scoring table: BLOSUM62
Gapcp 10.0 , Gapext 0.5

Searched: 562222 seqs, 172934929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPREMBL_19;*

1: sp_archea:*

2: sp_bacteria:*

3: sp_fungi:*

4: sp_human:*

5: sp_invertebrate:*

6: sp_mammal:*

7: sp_mhc:*

8: sp_organelle:*

9: sp_phage:*

10: sp_plant:*

11: sp_rabbit:*

12: sp_virus:*

13: sp_vertebrate:*

14: sp_unclassified:*

15: sp_rvirus:*

16: sp_bacteriap:*

17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	69	75.0	460	5 Q9GZF7
2	66	71.7	291	5 Q19470
3	63	68.5	1366	5 Q15177
4	62	67.4	326	2 Q9AGC6
5	61	66.3	410	2 Q9F691
6	60	65.2	358	12 Q9J3U4
7	60	65.2	566	11 Q9IV14
8	59	64.1	328	5 Q9U336
9	59	64.1	380	2 Q9F688
10	59	64.1	739	11 Q70575
11	59	64.1	751	11 Q9R1N9
12	59	64.1	1017	11 Q9K41
13	58	63.0	434	2 Q9ACN4
14	57	62.0	1051	5 Q26055
15	57	62.0	1352	13 Q9OY50
16	57	62.0	1419	11 Q63123

ALIGNMENTS

RESULT 1

ID	Q9GZF7	PRELIMINARY:	PRY:	460 AA.
AC	Q9GZF7;			
DT	01-MAR-2001 (TREMBLrel. 16, Created)			
DT	01-MAR-2001 (TREMBLrel. 16, Last sequence update)			
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)			
DE	HYPOTHETICAL 41.0 KDA PROTEIN.			
GN	C18H7_3.			
OS	Caenorhabditis elegans.			
OC	Eukaryota; Metazoa; Chromadorea; Rhabditida; Rhabditoidea;			
OC	Rhabditidae; Pelodermatae; Caenorhabditis.			
OX	NCBI TaxID:6239;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-BRISTOL N2;			
RX	MEDLINE=99069613; Pubmed=9951916;			
RA	None;			
RT	Investigating biology: The C. elegans Sequencing Consortium. ";			
RT	Science 282:2012-2016(1998).			
RL	{2}			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-BRISTOL N2;			
RA	Tin-Wollam A., Fronick W.;			
RT	The sequence of C. elegans cosmid C18H7. ";			
RL	Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.			
RN	{3}			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-BRISTOL N2;			
RA	Waterton R.;			
RT	"Direct Submission";			
RL	Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AF061607; AAF98809.1. -.			
DR	IntPro; IPR00087; Collagen.			
DR	InterPro; IPR002486; Collagene_N.			
DR	PFAM; PF01331; Collagen; 2.			
DR	PFAM; PF01184; Col_cuticule_N; 1.			
KW	Hypothetical protein.			

SQ	SEQUENCE	460 AA;	41016 MW;	BF99CC80770E202 CRC64;	RA	de Wet W., Bernard M., Benson-Chanda V., Chu M., Dickson L., Weil D.,
Query Match	75.0%	Score 69;	DB 5;	Length 460;	RA	Ramirez F.;
Best Local Similarity	80.0%	Pred. No.	0.021;		RT	"Organization of the human pro-alpha2(I) collagen gene.";
Matches	12;	Conservative	1;	Mismatches	2;	J. Biol. Chem. 262:16032-16036(1987).
QY	1	PAGPPGPNKGDKGVG	15		RL	[3]
Db	237	PAGPPGPNKGDKGEAG	251		RP	SEQUENCE OF 958-1351 FROM N.A.
RESULT 2					RX	MEDLINE=99304220; PubMed=2364107;
ID	019470	PRELIMINARY;	PRT;	291 AA.	RA	Makeila J.K., Vuorio T., Vuorio E.;
AC	Q19470;				RT	"Growth-dependent modulation of type I collagen production and mRNA
DT	01-NOV-1996	(TREMBLrel. 01, created)			RL	levels in cultured human skin fibroblasts.";
DT	01-NOV-1996	(TREMBLrel. 01, last sequence update)			RT	Biochim. Biophys. Acta 1049:171-176(1990).
DT	01-DEC-2001	(TREMBLrel. 19, last annotation update)			RL	[4]
DE	F15A2.1	PROTEIN.			RP	SEQUENCE FROM N.A.
GN	F15A2.1				RA	Dalgleish R.;
OS	Caenorhabditis elegans.				RL	Submitted (JUL-1996) to the EMBL/GenBank/DDBJ databases.
OC	Bukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;				RN	[5]
OC	Rhabdite; peloderrinae; Caenorhabditis.				RP	SEQUENCE FROM N.A.
OX	NCBI_TAXID=6239;				RX	MEDLINE=97169339; PubMed=9016532;
RN	[1]	SEQUENCE FROM N.A.			RA	Dalgleish R.;
RP	Gregory J.;				RT	"The human type I collagen mutation database.";
RA	submitted (MAR-1996) to the EMBL/GenBank/DDBJ databases.				RL	Nucleic Acids Res. 25:181-187(1997).
RL	[2]	SEQUENCE FROM N.A.			DR	EMBL; 274616; CAA98169.1; -.
RP	MEDLINE=99069613; PubMed=9851916;				DR	InterPro; IPR000077; Collagen.
RA	none;				DR	InterPro; IPR00885; Fib.collagen.C.
RT	"Genome sequence of the nematode <i>C.elegans</i> : A platform for investigating biology.";				DR	Pfam; PF01410; COLFI; 1.
RT	Science 282:2012-2018(1998).				DR	Pfam; PF01391; Collagen; 18.
DR	EMBL; Z71207; CAA94128.1; -.				DR	ProDom; P002018; Fib.collagen.C; 1.
DR	InterPro; IPR002486; Collagen.				DR	SMART; SM00038; COLFI; 1.
DR	Pfam; PF01391; Col1agén; 2.				KW	Signal; Collagen.
DR	Pfam; PF01484; Col1cuticle_N.				FT	Signal 1 22 POTENTIAL.
SO	SEQUENCE FROM N.A.				FT	SEQUENCE 1366 AA; 129337 MW; 5796859E6E50286C CRC64;
RX	Best Local Similarity 80.0%; Pred. No. 0.036;	Score 66;	DB 5;	Length 291;	SO	Query Match 68.5%; Score 63;
Matches	12;	Conservative	0;	Mismatches	ID	DB 4;
RA	0	Indels	3;	Indels	AC	Length 1366;
OY	1	PAGPPGPNKGDKGVG	15	0;	Q9AGC6	Best Local Similarity 73.3%; Pred. No. 0.52;
Db	183	PAGPAGPEKGDKGPVG	197	0;	Q9AGC6	Matches 11; Conservative 1; Mismatches 3;
RESULT 3					DT	Indels 0; Gaps 0;
Q15177	PRELIMINARY;	PRT;	1356 AA.		DT	Indels 0; Gaps 0;
ID	Q15177				DT	Indels 0; Gaps 0;
AC	Q15177;				DT	Indels 0; Gaps 0;
DT	01-NOV-1996	(TREMBLrel. 01, created)			DT	Indels 0; Gaps 0;
DT	01-NOV-1996	(TREMBLrel. 01, last sequence update)			DT	Indels 0; Gaps 0;
DT	01-DEC-2001	(TREMBLrel. 19, last annotation update)			DT	Indels 0; Gaps 0;
DE	PREPRO-ALPHA2(1) COLLAGEN PRECURSOR.				DE	Indels 0; Gaps 0;
OS	Homo sapiens (Human).				OS	COLLAGEN-LIKE PROTEIN B (FRAGMENT).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				OC	Streptococcus pyogenes.
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OX	NCBI_TAXID=9605;				OC	Streptococcus.
RN	[1]	SEQUENCE OF 1-765 FROM N.A.			OX	NCBI_TAXID=1314;
RX	MEDLINE=88339824; PubMed=3424913;				RN	[1]
RX	Kuivaniemi H., Tromp G., Chu M., Prockop D.J.;				RP	SEQUENCE FROM N.A.
RT	*Structure of a full-length cDNA clone for the proalpha2(I) chain of human type I procollagen.;				RC	STRAN=AP12;
RT	Biochem. J. 252:633-640(1988).				RA	Rasmussen M., Björck L.;
RL	[2]	PARTIAL SEQUENCE FROM N.A.			RT	"Phase-variation of a novel collagen-like surface protein in Streptococcus pyogenes."
RX	MEDLINE=88339824; PubMed=3424913;				RL	Submitted (JAN-2001) to the EMBL/GenBank/DDBJ databases.
RX	Kuivaniemi H., Tromp G., Chu M., Prockop D.J.;				DR	EMBL; AF36811; AAK0077.1; -.
RT	*Structure of a full-length cDNA clone for the proalpha2(I) chain of human type I procollagen.;				DR	InterPro; IPR00087; Collagen.
RL	Biochem. J. 252:633-640(1988).				DR	Pfam; PF01391; Collagen; 3.
RN	[2]	PARTIAL SEQUENCE FROM N.A.			FT	NON_TER 1 1
RX	MEDLINE=88058962; PubMed=2824475;				FT	NON_TER 326 326 AA; 33383 MW; BEE52296FF2445B CRC64;
Query Match	67.4%	Score 62;	DB 2;	Length 326;	SO	Query Match 67.4%; Score 62;
Best Local Similarity	73.3%	Pred. No.	0.16;		ID	Best Local Similarity 73.3%; Pred. No. 0.16;
Matches	11;	Conservative	1;	Mismatches	AC	Matches 11; Conservative 1; Mismatches 3;
RA	0	Indels	0;	Gaps	Q9AGC6	Indels 0; Gaps 0;
OY	1	PAGPPGPNKGDKGVG	15	0;	Q9AGC6	Indels 0; Gaps 0;

Db 126 PAGPVGPKGDGEG 140

RESULT 5

ID 09F691 PRELIMINARY; PRT; 410 AA.

AC 09F691; DT 01-MAR-2001 (TREMBrel. 16, Created)

DT 01-MAR-2001 (TREMBrel. 16, Last sequence update)

DT 01-JUN-2001 (TREMBrel. 17, Last annotation update)

DE COLLAGEN-LIKE SURFACE PROTEIN (FRAGMENT).

OS Streptococcus pyogenes.

Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;

OC Streptococcus.

OX NCBI_TAXID=1314;

RN [1]

SEQUENCE FROM N.A.

RC MEDLINE=20490571; PubMed=11035747;

RA Rasmussen M.; Eden A.; Björck L.;

RT "SCIA, a novel collagen-like surface protein of streptococcus

RL Infect. Immun. 68:6370-6377(2000).

EMBL: AF296332; AAQ630212_1; -.

DR InterPro; IPR00087; Collagen.

DR InterPro; IPR002965; P_rich_extensn.

DR Pfam; PF01391; Collagen; 2.

DR PRINTS; PRO1217; PRICHEXTENSN.

FT NON_TER 1

FT NON_TER 410 410

SQ SEQUENCE 410 AA; 42464 MW; D45596/0E7ECDA0F CRC64;

Query Match 66.3%; Score 61; DB 2; Length 410;

Matches 11; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 PAGGPGPNCKGKVG 15

Db 118 PAGGPAGPEKGPKAG 132

RESULT 6

ID 09J3U4 PRELIMINARY; PRT; 358 AA.

AC 09J3U4; DT 01-OCT-2000 (TREMBrel. 15, Created)

DT 01-OCT-2000 (TREMBrel. 15, Last sequence update)

DT 01-JUN-2001 (TREMBrel. 17, Last annotation update)

DE COLLAGEN-LIKE PROTEIN 1.

OS Ectocarpus siliculosus virus.

OC Virtues; dsDNA viruses, no RNA stage; Phycodnaviridae.

OX NCBI_TAXID=37665;

RN [1]

SEQUENCE FROM N.A.

RC STRAIN=EV-1;

RC MEDLINE=20192171; PubMed=10725207;

RA Delaroque N.; Wolf S.; Müller D.G.; Klipplers R.;

RT "Characterization and immunolocalization of major structural proteins

RT in the brown algal virus EV-1.";

RT Virology 269:148-155(2000).

DR EMBL: AF204951; AA#28321_1; -.

DR InterPro; IPR00087; Collagen.

DR Pfam; PF01391; Collagen; 3.

SQ SEQUENCE 358 AA; 33690 MW; A9A07CC077AA025A CRC64;

Query Match 65.2%; Score 60; DB 12; Length 358;

Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 PAGGPGPNCKGKVG 15

Db 264 PAGGPAGPEKGPKAG 278

RESULT 7

ID 09IVL4 PRELIMINARY; PRT; 566 AA.

AC 09IVL4; DT 01-DEC-2001 (TREMBrel. 19, Created)

DT 01-DEC-2001 (TREMBrel. 19, Last sequence update)

DE UNKNOWN (PROTEIN FOR IMAGE:3481648) (FRAGMENT).

OS Mus musculus (Mouse).

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TAXID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=BREAST TUMOR;

RA Strausberg R.;

FT Submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; BC012430; AAM12430_1; -.

FT NON_TER 1

SQ SEQUENCE 566 AA; 56693 MW; 2E845DDDFC7BBB0D5 CRC64;

Query Match 65.2%; Score 60; DB 11; Length 566;

Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 PAGCPGPGPKGKVG 15

Db 255 PAGPSGPVKGDGRSG 269

RESULT 8

ID 09U336 PRELIMINARY; PRT; 328 AA.

AC 09U336; DT 01-MAY-2000 (TREMBrel. 13, Created)

DT 01-MAY-2000 (TREMBrel. 13, Last sequence update)

DT 01-JUN-2001 (TREMBrel. 17, Last annotation update)

DE W07A12.5 PROTEIN.

GN W07A12.5.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nemata; Chromadorea; Rhabditida; Rhabditoidae.

OC Rhabditidae; Peloderinae; Caenorhabditis.

OX NCBI_TAXID=6239;

RN [1]

RP SEQUENCE FROM N.A.

RA Thomas K.;

RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RA MEDLINE=99069613; PubMed=9851916;

RA note;

"Genome sequence of the nematode C. elegans: A platform for

RT investigating biology.";

RT Science 282:2012-2018(1998).

DR EMBL; Z66320; CAB63328_1; -.

DR InterPro; IPR00087; Collagen.

DR Pfam; PF01391; Collagen; 2.

SQ SEQUENCE 328 AA; 33720 MW; 6178EEF08FB884AC5 CRC64;

Query Match 64.1%; Score 59; DB 5; Length 328;

Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 1 PAGCPGPGPKGKVG 15

Db 221 PVGPGPGPKGDKGSIG 235

RESULT 9

ID 09F688 PRELIMINARY; PRT; 380 AA.

AC 09F688;

DT 01-MAR-2001 (TREMBLrel. 16; Created)
 DT 01-MAR-2001 (TREMBLrel. 15; Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17; Last annotation update)

DE COLLAGEN-LIKE SURFACE PROTEIN (FRAGMENT).

OS streptococcus progenes.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;

OC streptococcus.

OX NCBI_TAXID=1314;
 RN [1]

RP SEQUENCE FROM N A.

RC STRAIN=KTL9;

RX MEDLINE=20-90571; PubMed=11035747;

RA Rasmussen M., Eden A., Björck L.;
 RT "SCLA, a novel collagen-like surface protein of streptococcus
 pyogenes.";
 RL Infect. Immun. 68:6370-6377 (2000).

DR EMBL: AF296335; AAC0215.1; -.
 InterPro: IPR000087; Collagen.

DR Pfam: PF01391; Collagen; 3.

FT NON_TER 380 AA; 380 MW; B6D283C3C7B717A4 CRC64;

SQ SEQUENCE 380 AA; 39224 MW; B6D283C3C7B717A4 CRC64;

Query Match 64.1%; Score 59; DB 2; Length 380;
 Best Local Similarity 66.7%; Pred. No. 0 53; 1; Mismatches 4; Indels 0; Gaps 0;
 Matches 10; Conservative 10;

Oy 1 PAGPWGNGKDGKV 15
 1 | || | :|| |
 142 POGPVGPKGEDOKKAG 156

RESULT 10

OY 070575 PRELIMINARY; PRT; 739 AA.

ID 070575; TREPMLREL. 07, Created)
 DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)

DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

DE COLLAGEN TYPE XIII ALPHA-1 CHAIN.

GN COL13A1.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

OX NCBI_TaxID=10090;
 RN [1]

SEQUENCE FROM N A.

RC TISSUE=GUT;

RX MEDLINE=98288296; PubMed=9624150;

RA Hagg P., Rahn M., Huhtaniemi T., Vaisanen T., Tamminen M.,
 RA Pihlajaniemi T.;
 RT "Type XIII collagen is identified as a plasma membrane protein.";
 RL J. Biol. Chem. 273:15590-15597(1998).

DR EMBL: U310292; AAC24314.1; -.
 DR MGI: MGI:1277201; Col13a1.

DR InterPro: IPR000087; Collagen.

DR Pfam: PF01391; Collagen; 9.

DR SEQUENCE 739 AA; F13951061381F017 CRC64;

Query Match 64.1%; Score 59; DB 11; Length 739;
 Best Local Similarity 73.3%; Pred. No. 1.1; Mismatches 4; Indels 0; Gaps 0;
 Matches 11; Conservative 11;

Oy 1 PAGPWGNGKDGKV 15
 1 | || | :|| |
 Db 514 PAGPQGPQKDGPPG 528

RESULT 11

OY 09R1N9 PRELIMINARY; PRT; 751 AA.

ID 09R1N9; TREPMLREL. 13, Created)

RESULT 12

OY 099K41 PRELIMINARY; PRT; 1017 AA.

ID 099K41; TREPMLREL. 17, Created)
 AC 099K41;
 DT 01-JUN-2001 (TREMBLrel. 17; Last sequence update)

DT 01-JUN-2001 (TREMBLrel. 19, Last annotation update)

DE SIMILAR TO ELASTIN MICROFIBRIL INTERFACE LOCATED PROTEIN.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

OX NCBI_TaxID=10090;
 RN [1]

DT 01-MAY-2000 (TREMBLrel. 13; Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

DE TYPE XIII COLLAGEN.

GN COL13A1 OR COL13A1.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;
 RN [1]

RP SEQUENCE FROM N A.

RC STRAIN=KTL9;

RX MEDLINE=99357014; PubMed=10429945;

RA Palotie A., Beier D., Piha-Jananiemi T.;
 RT "Complete exon-intron organization and chromosomal location of the
 gene for mouse type XIII collagen (coll13a1) and comparison with its
 human homologue.";
 RL Matrix Biol. 18:261-274 (1999).

DR EMBL: AF063693; AAD50327.1; -.

DR EMBL: AF063666; AAD50327.1; JOINED.

DR EMBL: AF063667; AAD50327.1; JOINED.

DR EMBL: AF063668; AAD50327.1; JOINED.

DR EMBL: AF063669; AAD50327.1; JOINED.

DR EMBL: AF063670; AAD50327.1; JOINED.

DR EMBL: AF063672; AAD50327.1; JOINED.

DR EMBL: AF063673; AAD50327.1; JOINED.

DR EMBL: AF063674; AAD50327.1; JOINED.

DR EMBL: AF063675; AAD50327.1; JOINED.

DR EMBL: AF063669; AAD50327.1; JOINED.

DR EMBL: AF063677; AAD50327.1; JOINED.

DR EMBL: AF063678; AAD50327.1; JOINED.

DR EMBL: AF063679; AAD50327.1; JOINED.

DR EMBL: AF063680; AAD50327.1; JOINED.

DR EMBL: AF063681; AAD50327.1; JOINED.

DR EMBL: AF063682; AAD50327.1; JOINED.

DR EMBL: AF063683; AAD50327.1; JOINED.

DR EMBL: AF063684; AAD50327.1; JOINED.

DR EMBL: AF063685; AAD50327.1; JOINED.

DR EMBL: AF063686; AAD50327.1; JOINED.

DR EMBL: AF063687; AAD50327.1; JOINED.

DR EMBL: AF063688; AAD50327.1; JOINED.

DR EMBL: AF063689; AAD50327.1; JOINED.

DR EMBL: AF063690; AAD50327.1; JOINED.

DR EMBL: AF063691; AAD50327.1; JOINED.

DR EMBL: AF063692; AAD50327.1; JOINED.

DR EMBL: AF063693; AAD50327.1; JOINED.

DR EMBL: AF063694; AAD50327.1; JOINED.

DR EMBL: AF063695; AAD50327.1; JOINED.

DR EMBL: AF063696; AAD50327.1; JOINED.

DR EMBL: AF063697; AAD50327.1; JOINED.

DR EMBL: AF063698; AAD50327.1; JOINED.

DR EMBL: AF063699; AAD50327.1; JOINED.

DR EMBL: AF063680; AAD50327.1; JOINED.

DR EMBL: AF063681; AAD50327.1; JOINED.

DR EMBL: AF063682; AAD50327.1; JOINED.

DR EMBL: AF063683; AAD50327.1; JOINED.

DR EMBL: AF063684; AAD50327.1; JOINED.

DR EMBL: AF063685; AAD50327.1; JOINED.

DR EMBL: AF063686; AAD50327.1; JOINED.

DR EMBL: AF063687; AAD50327.1; JOINED.

DR EMBL: AF063688; AAD50327.1; JOINED.

DR EMBL: AF063689; AAD50327.1; JOINED.

DR EMBL: AF063690; AAD50327.1; JOINED.

DR EMBL: AF063691; AAD50327.1; JOINED.

DR EMBL: AF063692; AAD50327.1; JOINED.

DR EMBL: AF063693; AAD50327.1; JOINED.

DR EMBL: AF063694; AAD50327.1; JOINED.

DR EMBL: AF063695; AAD50327.1; JOINED.

DR EMBL: AF063696; AAD50327.1; JOINED.

DR EMBL: AF063697; AAD50327.1; JOINED.

DR EMBL: AF063698; AAD50327.1; JOINED.

DR EMBL: AF063699; AAD50327.1; JOINED.

DR EMBL: AF063680; AAD50327.1; JOINED.

DR EMBL: AF063681; AAD50327.1; JOINED.

DR EMBL: AF063682; AAD50327.1; JOINED.

DR EMBL: AF063683; AAD50327.1; JOINED.

DR EMBL: AF063684; AAD50327.1; JOINED.

DR EMBL: AF063685; AAD50327.1; JOINED.

DR EMBL: AF063686; AAD50327.1; JOINED.

DR EMBL: AF063687; AAD50327.1; JOINED.

DR EMBL: AF063688; AAD50327.1; JOINED.

DR EMBL: AF063689; AAD50327.1; JOINED.

DR EMBL: AF063690; AAD50327.1; JOINED.

DR EMBL: AF063691; AAD50327.1; JOINED.

DR EMBL: AF063692; AAD50327.1; JOINED.

DR EMBL: AF063693; AAD50327.1; JOINED.

DR EMBL: AF063694; AAD50327.1; JOINED.

DR EMBL: AF063695; AAD50327.1; JOINED.

DR EMBL: AF063696; AAD50327.1; JOINED.

DR EMBL: AF063697; AAD50327.1; JOINED.

DR EMBL: AF063698; AAD50327.1; JOINED.

DR EMBL: AF063699; AAD50327.1; JOINED.

DR EMBL: AF063680; AAD50327.1; JOINED.

DR EMBL: AF063681; AAD50327.1; JOINED.

DR EMBL: AF063682; AAD50327.1; JOINED.

DR EMBL: AF063683; AAD50327.1; JOINED.

DR EMBL: AF063684; AAD50327.1; JOINED.

DR EMBL: AF063685; AAD50327.1; JOINED.

DR EMBL: AF063686; AAD50327.1; JOINED.

DR EMBL: AF063687; AAD50327.1; JOINED.

DR EMBL: AF063688; AAD50327.1; JOINED.

DR EMBL: AF063689; AAD50327.1; JOINED.

DR EMBL: AF063690; AAD50327.1; JOINED.

DR EMBL: AF063691; AAD50327.1; JOINED.

DR EMBL: AF063692; AAD50327.1; JOINED.

DR EMBL: AF063693; AAD50327.1; JOINED.

DR EMBL: AF063694; AAD50327.1; JOINED.

DR EMBL: AF063695; AAD50327.1; JOINED.

DR EMBL: AF063696; AAD50327.1; JOINED.

DR EMBL: AF063697; AAD50327.1; JOINED.

DR EMBL: AF063698; AAD50327.1; JOINED.

DR EMBL: AF063699; AAD50327.1; JOINED.

DR EMBL: AF063680; AAD50327.1; JOINED.

DR EMBL: AF063681; AAD50327.1; JOINED.

DR EMBL: AF063682; AAD50327.1; JOINED.

DR EMBL: AF063683; AAD50327.1; JOINED.

DR EMBL: AF063684; AAD50327.1; JOINED.

DR EMBL: AF063685; AAD50327.1; JOINED.

DR EMBL: AF063686; AAD50327.1; JOINED.

DR EMBL: AF063687; AAD50327.1; JOINED.

DR EMBL: AF063688; AAD50327.1; JOINED.

DR EMBL: AF063689; AAD50327.1; JOINED.

DR EMBL: AF063690; AAD50327.1; JOINED.

DR EMBL: AF063691; AAD50327.1; JOINED.

DR EMBL: AF063692; AAD50327.1; JOINED.

DR EMBL: AF063693; AAD50327.1; JOINED.

DR EMBL: AF063694; AAD50327.1; JOINED.

DR EMBL: AF063695; AAD50327.1; JOINED.

DR EMBL: AF063696; AAD50327.1; JOINED.

DR EMBL: AF063697; AAD50327.1; JOINED.

DR EMBL: AF063698; AAD50327.1; JOINED.

DR EMBL: AF063699; AAD50327.1; JOINED.

DR EMBL: AF063680; AAD50327.1; JOINED.

DR EMBL: AF063681; AAD50327.1; JOINED.

DR EMBL: AF063682; AAD50327.1; JOINED.

DR EMBL: AF063683; AAD50327.1; JOINED.

DR EMBL: AF063684; AAD50327.1; JOINED.

DR EMBL: AF063685; AAD50327.1; JOINED.

DR EMBL: AF063686; AAD50327.1; JOINED.

DR EMBL: AF063687; AAD50327.1; JOINED.

DR EMBL: AF063688; AAD50327.1; JOINED.

DR EMBL: AF063689; AAD50327.1; JOINED.

DR EMBL: AF063690; AAD50327.1; JOINED.

DR EMBL: AF063691; AAD50327.1; JOINED.

DR EMBL: AF063692; AAD50327.1; JOINED.

DR EMBL: AF063693; AAD50327.1; JOINED.

DR EMBL: AF063694; AAD50327.1; JOINED.

DR EMBL: AF063695; AAD50327.1; JOINED.

DR EMBL: AF063696; AAD50327.1; JOINED.

DR EMBL: AF063697; AAD50327.1; JOINED.

DR EMBL: AF063698; AAD50327.1; JOINED.

DR EMBL: AF063699; AAD50327.1; JOINED.

DR EMBL: AF063680; AAD50327.1; JOINED.

DR EMBL: AF063681; AAD50327.1; JOINED.

DR EMBL: AF063682; AAD50327.1; JOINED.

DR EMBL: AF063683; AAD50327.1; JOINED.

DR EMBL: AF063684; AAD50327.1; JOINED.

DR EMBL: AF063685; AAD50327.1; JOINED.

DR EMBL: AF063686; AAD50327.1; JOINED.

DR EMBL: AF063687; AAD50327.1; JOINED.

DR EMBL: AF063688; AAD50327.1; JOINED.

DR EMBL: AF063689; AAD50327.1; JOINED.

DR EMBL: AF063690; AAD50327.1; JOINED.

DR EMBL: AF063691; AAD50327.1; JOINED.

DR EMBL: AF063692; AAD50327.1; JOINED.

DR EMBL: AF063693; AAD50327.1; JOINED.

DR EMBL: AF063694; AAD50327.1; JOINED.

DR EMBL: AF063695; AAD50327.1; JOINED.

DR EMBL: AF063696; AAD50327.1; JOINED.

DR EMBL: AF063697; AAD50327.1; JOINED.

DR EMBL: AF063698; AAD50327.1; JOINED.

DR EMBL: AF063699; AAD50327.1; JOINED.

DR EMBL: AF063680; AAD50327.1; JOINED.

DR EMBL: AF063681; AAD50327.1; JOINED.

DR EMBL: AF063682; AAD50327.1; JOINED.

DR EMBL: AF063683; AAD50327.1; JOINED.

DR EMBL: AF063684; AAD50327.1; JOINED.

DR EMBL: AF063685; AAD50327.1; JOINED.

DR EMBL: AF063686; AAD50327.1; JOINED.

DR EMBL: AF063687; AAD50327.1; JOINED.

DR EMBL: AF063688; AAD50327.1; JOINED.

DR EMBL: AF063689; AAD50327.1; JOINED.

DR EMBL: AF063690; AAD50327.1; JOINED.

DR EMBL: AF063691; AAD50327.1; JOINED.

DR EMBL: AF063692; AAD50327.1; JOINED.

DR EMBL: AF063693; AAD50327.1; JOINED.

DR EMBL: AF063694; AAD50327.1; JOINED.

DR EMBL: AF063695; AAD50327.1; JOINED.

DR EMBL: AF063696; AAD50327.1; JOINED.

DR EMBL: AF063697; AAD50327.1; JOINED.

DR EMBL: AF063698; AAD50327.1; JOINED.

DR EMBL: AF063699; AAD50327.1; JOINED.

DR EMBL: AF063680; AAD50327.1; JOINED.

DR EMBL: AF063681; AAD50327.1; JOINED.

DR EMBL: AF063682; AAD50327.1; JOINED.

DR EMBL: AF063683; AAD50327.1; JOINED.

DR EMBL: AF063684; AAD50327.1; JOINED.

DR EMBL: AF063685; AAD50327.1; JOINED.

DR EMBL: AF063686; AAD50327.1; JOINED.

DR EMBL: AF063687; AAD50327.1; JOINED.

DR EMBL: AF063688; AAD50327.1; JOINED.

DR EMBL: AF063689; AAD50327.1; JOINED.

DR EMBL: AF063690; AAD50327.1; JOINED.

DR EMBL: AF063691; AAD50327.1; JOINED.

DR EMBL: AF063692; AAD50327.1; JOINED.

DR EMBL: AF063693; AAD50327.1; JOINED.

DR EMBL: AF063694; AAD50327.1; JOINED.

DR EMBL: AF063695; AAD50327.1; JOINED.

DR EMBL: AF063696; AAD50327.1; JOINED.

DR EMBL: AF063697; AAD50327.1; JOINED.

DR EMBL: AF063698; AAD50327.1; JOINED.

DR EMBL: AF063699; AAD50327.1; JOINED.

DR EMBL: AF063680; AAD50327.1; JOINED.

DR EMBL: AF063681; AAD50327.1; JOINED.

DR EMBL: AF063682; AAD50327.1; JOINED.

DR EMBL: AF063683; AAD50327.1; JOINED.

DR EMBL: AF063684; AAD50327.1; JOINED.

DR EMBL: AF063685; AAD50327.1; JOINED.

DR EMBL: AF063686; AAD50327.1; JOINED.

DR EMBL: AF063687; AAD50327.1; JOINED.

DR EMBL: AF063688; AAD50327.1; JOINED.

DR EMBL: AF063689; AAD50327.1; JOINED.

DR EMBL: AF063690; AAD50327.1; JOINED.

DR EMBL: AF063691; AAD50327.1; JOINED.

DR EMBL: AF063692; AAD50327.1; JOINED.

DR EMBL: AF063693; AAD50327.1; JOINED.

DR EMBL: AF063694; AAD50327.1; JOINED.

DR EMBL: AF063695; AAD50327.1; JOINED.

DR EMBL: AF063696; AAD50327.1; JOINED.

DR EMBL: AF063697; AAD50327.1; JOINED.

DR EMBL: AF063698; AAD50327.1; JOINED.

DR EMBL: AF063699; AAD50327.1; JOINED.

DR EMBL: AF063680; AAD50327.1; JOINED.

DR EMBL: AF063681; AAD50327.1; JOINED.

DR EMBL: AF063682; AAD50327.1; JOINED.

DR EMBL: AF063683; AAD50327.1; JOINED.

DR EMBL: AF063684; AAD50327.1; JOINED.

DR EMBL: AF063685; AAD50327.1; JOINED.

DR EMBL: AF063686; AAD50327.1; JOINED.

DR EMBL: AF063687; AAD50327.1; JOINED.

DR EMBL: AF063688; AAD50327.1; JOINED.

DR EMBL: AF063689; AAD50327.1; JOINED.

DR EMBL: AF063690; AAD50327.1; JOINED.

DR EMBL: AF063691; AAD50327.1; JOINED.

DR EMBL: AF063692; AAD50327.1; JOINED.

DR EMBL: AF063693; AAD50327.1; JOINED.

DR EMBL: AF063694; AAD50327.1; JOINED.

DR EMBL: AF063695; AAD50327.1; JOINED.

DR EMBL: AF063696; AAD50327.1; JOINED.

DR EMBL: AF063697; AAD50327.1; JOINED.

DR EMBL: AF063698; AAD50327.1; JOINED.

DR EMBL: AF063699; AAD50327.1; JOINED.

DR EMBL: AF063680; AAD50327.1; JOINED.

DR EMBL: AF063681; AAD50327.1; JOINED.

DR EMBL: AF063682; AAD50327.1; JOINED.

DR EMBL: AF063683; AAD50327.1; JOINED.

DR EMBL: AF063684; AAD50327.1; JOINED.

DR EMBL: AF063685; AAD50327.1; JOINED.

DR EMBL: AF063686; AAD50327.1; JOINED.

DR EMBL: AF063687; AAD50327.1; JOINED.

DR EMBL: AF063688; AAD50327.1; JOINED.

DR EMBL: AF063689; AAD50327.1; JOINED.

DR EMBL: AF063690; AAD50327.1; JOINED.

DR EMBL: AF063691; AAD50327.1; JOINED.

DR EMBL: AF063692; AAD50327.1; JOINED.

DR EMBL: AF063693; AAD50327.1; JOINED.

DR EMBL: AF063694; AAD50327.1; JOINED.

DR EMBL: AF063695; AAD50327.1; JOINED.

DR EMBL: AF063696; AAD50327.1; JOINED.

DR EMBL: AF063697; AAD50327.1; JOINED.

DR EMBL: AF063698; AAD50327.1; JOINED.

DR EMBL: AF063699; AAD50327.1; JOINED.

DR EMBL: AF063680; AAD50327.1; JOINED.

DR EMBL: AF063681; AAD50327.1; JOINED.

DR EMBL: AF063682; AAD50327.1; JOINED.

DR EMBL: AF063683; AAD50327.1; JOINED.

DR EMBL: AF063684; AAD50327.1; JOINED.

DR EMBL: AF063685; AAD50327.1; JOINED.

DR EMBL: AF063686; AAD50327.1; JOINED.

DR EMBL: AF063687; AAD50327.1; JOINED.

DR EMBL: AF063688; AAD50327.1; JOINED.

DR EMBL: AF063689; AAD50327.1; JOINED.

DR EMBL: AF063690; AAD50327.1; JOINED.

DR EMBL: AF063691; AAD50327.1; JOINED.

DR EMBL: AF063692; AAD50327.1; JOINED.

DR EMBL: AF063693; AAD50327.1; JOINED.

DR EMBL: AF063694; AAD50327.1; JOINED.

DR EMBL: AF063695; AAD50327.1; JOINED.

DR EMBL: AF063696; AAD50327.1; JOINED.

DR EMBL: AF063697; AAD50327.1; JOINED.

DR EMBL: AF063698; AAD50327.1; JOINED.

DR EMBL: AF063699; AAD50327.1; JOINED.

DR EMBL: AF063680; AAD50327.1; JOINED.

DR EMBL: AF063681; AAD50327.1; JOINED.

DR EMBL: AF063682; AAD50327.1; JOINED.

DR EMBL: AF063683; AAD50327.1; JOINED.

DR EMBL: AF063684; AAD50327.1; JOINED.

DR EMBL: AF063685; AAD50327.1; JOINED.

DR EMBL: AF063686; AAD50327.1; JOINED.

DR EMBL: AF063687; AAD50327.1; JOINED.

DR EMBL: AF063688; AAD50327.1; JOINED.

DR EMBL: AF063689; AAD50327.1; JOINED.

DR EMBL: AF063690; AAD50327.1; JOINED.

DR EMBL: AF063691; AAD50327.1; JOINED.

DR EMBL: AF063692; AAD50327.1; JOINED.

DR EMBL: AF063693; AAD50327.1; JOINED.

DR EMBL: AF063694; AAD50327.1; JOINED.

DR EMBL: AF063695; AAD50327.1; JOINED.

DR EMBL: AF063696; AAD50327.1; JOINED.

DR EMBL: AF063697; AAD50327.1; JOINED.

DR EMBL: AF063698; AAD50327.1; JOINED.

DR EMBL: AF063699; AAD50327.1; JOINED.

DR EMBL: AF063680; AAD50327.1; JOINED.

DR EMBL: AF063681; AAD50327.1; JOINED.

DR EMBL: AF063682; AAD50327.1; JOINED.

DR EMBL: AF063683; AAD50327.1; JOINED.

DR EMBL: AF063684; AAD50327.1; JOINED.

DR EMBL: AF063685; AAD50327.1; JOINED.

DR EMBL: AF063686; AAD50327.1; JOINED.

DR EMBL: AF063687; AAD50327.1; JOINED.

DR EMBL: AF063688; AAD50327.1; JOINED.

DR EMBL: AF063689; AAD50327.1; JOINED.

DR EMBL: AF063690; AAD50327.1; JOINED.

DR EMBL: AF063691; AAD50327.1; JOINED.

DR EMBL: AF063692; AAD50327.1; JOINED.

DR EMBL: AF063693; AAD50327.1; JOINED.

DR EMBL: AF063694; AAD50327.1; JOINED.

DR EMBL: AF063695; AAD50327.1; JOINED.

DR EMBL: AF063696; AAD50327.1; JOINED.</

RP SEQUENCE FROM N.A.
 RC TISSUE=MAMMARY TUMOR. WAP-TGF ALPHA MODEL. 7 MONTHS OLD, GROSS
 RA TISSUE;
 RA Strausberg R.; Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 RL EMBL; BC005481; AAH05481; 1; .
 DR InterPro; IPR000073; C1Q.
 DR Pfam; PF00386; C1Q; 1.
 DR Pfam; PF01391; Collagen; 1.
 DR SMART; SM00110; C1Q; 1.
 SQ SEQUENCE 1017 AA; 107584 MW; 2EF903204DB9C80F CRC64;
 Query Match 64.1%; Score 59; DB 11; Length 1017;
 Best Local Similarity 73.3%; Pred. No. 1.5; Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 Qy 1 PAGPNCNGKDGKV 15
 Db 834 PAGPPGPGKDGQG 848

RESULT 13
 Q9ACN4 PRELIMINARY; PRT; 434 AA.
 ID Q9ACN4; 01-JUN-2001 ('T-EMBLrel. 17, created)
 AC Q9ACN4; 01-JUN-2001 ('T-EMBLrel. 17, last sequence update)
 DT 01-JUN-2001 ('T-EMBLrel. 19, last annotation update)
 DE SCLB PROTEIN.
 RN SEQUENCE FROM N.A.
 OS Streptococcus Pyogenes.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 OC Streptococcus.
 OC NCBI_TAXID=1314;
 RN [1] SEQUENCE FROM N.A.
 RP STRAIN=PI80;
 RX MEDLINE=21097281; PUBMED=11158359;
 RA Whatmore A.M.; Pyogenes sclb encodes a putative hypervariable surface
 RT "Streptococcus Pyogenes sclb encodes a putative repetitive structure";
 RT protein with a collagen-like repetitive structure.;
 RL Microbiology 147:419-429(2001).
 DR EMBL; AJ301807; CAC3776; 1; -.
 DR InterPro; IPR000087; Collagen.
 DR InterPro; IPR001899; Gram_Pos_anchor.
 DR Pfam; PF01391; Collagen; 4.
 DR Pfam; PF00746; Gram_pos_anchor; 1.
 DR PRINTS; PR00015; GPOSANCHOR.
 DR PROSITE; PS0033; GRAM_POS_ANCHORING; UNKNOWN_1.
 KW Transmembrane.
 SQ SEQUENCE 434 AA; 43885 MW; F77BB0DA881FD21 CRC64;

Query Match 63.0%; Score 58; DB 2; Length 434;
 Best Local Similarity 73.3%; Pred. No. 0.86; Matches 11; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 PAGPNCNGKDGKV 15
 Db 126 PAGPPGPGKDGQG 140

RESULT 14
 026055 PRELIMINARY; PRT; 1051 AA.
 ID 026055; 01-NOV-1996 ('T-EMBLrel. 01, Created)
 DT 01-NOV-1996 ('T-EMBLrel. 01, Last sequence update)
 DT 01-DEC-2001 ('T-EMBLrel. 19, Last annotation update)
 DE 2-ALPEA COLLAGEN (COL12-ALPHA) (FRAGMENT).
 OS Paracentrotus lividus (Common sea urchin).
 OC Eukaryota; Metazoa; Eublatherozoa; Echinodermata; Echinoidea; Echinacea; Echinidae; Echinidae;
 OC Paracentrotus; NCBI_TaxID=7656;
 RN [1]
 SEQUENCE FROM N.A.
 RX MEDLINE=90216744; Pubmed=2324112;
 RA D'Alessio M., Ramirez F., Suzuki H. R., Solursh M., Gambino R.;
 RA "Cloning of a fibrillar collagen gene expressed in the mesenchymal
 cells of the developing sea urchin embryo.";
 RL J. Biol. Chem. 265:7050-7054(1990).
 DR EMBL; J05422; AAA28440; 1; -.
 DR InterPro; IPR000087; Collagen.
 DR InterPro; IPR000885; Fib_Collagen_C.
 DR Pfam; PF01410; COLFI; 1.
 DR Pfam; PF01391; Collagen; 13.
 DR ProDom; PD00078; Fib_Collagen_C; 1.
 DR SMART; SM00038; COLFI; 1.
 KW Collagen.
 RP NON_TER
 FT CHAIN 1 1 786 POTENTIAL.
 SQ SEQUENCE 1051 AA; 101961 MW; 81295E33E82727C7 CRC64;

Query Match 62.0%; Score 57; DB 5; Length 1051;
 Best Local Similarity 66.7%; Pred. No. 3.1; Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 1 PAGPNCNGKDGKV 15
 Db 485 PAGPVGPNGRPGEDG 499

RESULT 15
 Q90YJ0 PRELIMINARY; PRT; 1352 AA.
 ID Q90YJ0; 01-DEC-2001 ('T-EMBLrel. 19, Created)
 AC Q90YJ0; 01-DEC-2001 ('T-EMBLrel. 19, last sequence update)
 DT 01-DEC-2001 ('T-EMBLrel. 19, last annotation update)
 DE PROCOLLAGEN TYPE I ALPHA 2 CHAIN.
 RN [1] SEQUENCE FROM N.A.
 OS Brachydanio rerio (Zebrafish) zebra danio.
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi;
 OC Actinopterygii; Neopteriogli; Teleostei; Ostariophysi;
 OC Cypriniformes; Cyprinidae; Danio.
 OC NCBI_TAXID=7955;
 RN [1] SEQUENCE FROM N.A.
 RA Morvan-Dubois G., Le Guellec D., Garrone R., Zylberberg L.,
 RA Bonnafont L.;
 RT "Phylogenetic analysis on vertebrates fibrillar collagen enlightens
 zebrafish a3(I) position and evidenced an evolutionary link between
 RT collagen alpha chains and Hox clusters.";
 RT Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
 RL [2] SEQUENCE FROM N.A.
 RN Morvan-Dubois G., Haftek Z., Crozet C., Garrone R., Le Guellec D.;
 RT "Structure and expression of the full length cDNA encoding zebrafish
 alpha 2 type I collagen.";
 RT Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
 RL EMBL; AJ318213; CAC51030; 1; -.
 DR Collagen.
 KW SEQUENCE 1352 AA; 127333 MW; 63D73C2CEAD44FE CRC64;

Query Match 62.0%; Score 57; DB 13; Length 1352;
 Best Local Similarity 73.3%; Pred. No. 4; Matches 11; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 PAGPNCNGKDGKV 15
 Db 1045 PAGPNPGKGDSNG 1059

RESULT 16
 063123

ID	Q03123	PRELIMINARY;	PRT;	1419 AA..;
DT	063123; Q63156;			
DT	01-NOV-1996 (TREMBlrel. 01, Created)			
DT	01-NOV-1996 (TREMBlrel. 01, Last sequence update)			
DT	01-JUN-2001 (TREMBlrel. 17, Last annotation update)			
DE	COLLAGEN ALPHA 1 TYPE II (T1 mRNA).			
T1	Rattus norvegicus (Rat);			
OS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
OX	NCBI_TAXID=10116;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=BONE FRACTURE CALLUS;			
RA	Urabe K.; Sarker G.; Bolander M.E.;			
RL	Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.			
RN	[2]			
RP	SEQUENCE OF 1372-1419 FROM N.A.			
RC	STRAIN=DA; TISSUE=Cartilage;			
RA	Wurtz T.; Brandstet C.; Lundmark C.; Christersson C.;			
RL	Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.			
RN	[3]			
RP	SEQUENCE OF 370-422 FROM N.A.			
RC	STRAIN=DA; TISSUE=Cartilage;			
RA	Michaelson E.; Malmstrom V.; Reis S.; Engstrom A.; Burkhardt H.;			
RT	Holmdahl R.;			
RT	"T cell recognition of carbohydrates on type II collagen.";			
RL	J. Exp. Med. 180:745-749(1994).			
DR	EMBL; L48440; AAA97801; -.			
DR	EMBL; AU224879; CAA12179; 1. -.			
DR	EMBL; X79816; CAA56213; 1. -.			
DR	InterPro; IPR000885; fib collagen_C.			
DR	InterPro; IPR001410; COLFI; 1.			
DR	Pfam; PF01391; Collagen_18.			
DR	Prodrom; PD002078; Fib collagen_C; 1.			
DR	SMART; SM00214; VWC; 1.			
DR	PROSITE; PS01208; VWF_C; 1.			
SQ	SEQUENCE 1419 AA; 134570 MW; B7C63B77819CE50B CRC64;			
Query Match	62.0%; Score 57; DB 11; Length 1419;			
Best Local Similarity	73.3%; Pred. No. 4.3;			
Matches	11; Conservative 0; Mismatches 0; Indels 4; Gaps 0; Gaps 0;			
Qy	1 PAGPWGPNGKDGKV 15			
Db	871 PAGPPGPASKDGPKG 885			
RESULT 18				
ID	Q62033	PRELIMINARY;	PRT;	1442 AA.
AC	Q62033;			
DT	062033; 01-NOV-1996 (TREMBlrel. 01, Created)			
DT	01-NOV-1996 (TREMBlrel. 01, Last sequence update)			
DT	01-DEC-2001 (TREMBlrel. 19, Last annotation update)			
DE	PRO-ALPHA-1 TYPE II COLLAGEN.			
GN	COL2A1.			
OS	Mus musculus (Mouse);			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TAXID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=C57/BLACK;			
RA	Metsaranta M.; Toman D.; de Crombrughe B.; Vuorio E.;			
RA	"Mouse type II collagen gene. Complete nucleotide sequence, exon structure, and alternative splicing.";			
RT	RT J. Biol. Chem. 266:16862-16869(1991).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=C57/BLACK;			
RA	Vuorio E.;			
RL	Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; M65161; AAA66102; 1. -.			
DR	MGD; MGI:88452; Col2a1.			
DR	InterPro; IPR00087; Collagen.			
DR	InterPro; IPR000885; fib collagen_C.			
DR	InterPro; IPR01007; VWF_C.			
DR	Pfam; PF0110; COLFI; 1.			
DR	Pfam; PF01391; Collagen_17.			
DR	Prodrom; PD002078; Fib collagen_C; 1.			
DR	SMART; SM00214; VWC; 1.			
DR	PROSITE; PS01208; VWF_C; 1.			
SQ	SEQUENCE 1442 AA; 137829 MW; F0E77C11BCAFA93B CRC64;			
Query Match	62.0%; Score 57; DB 11; Length 1442;			
Best Local Similarity	73.3%; Pred. No. 4.3;			
Matches	11; Conservative 0; Mismatches 4; Indels 0; Gaps 0; Gaps 0;			
Qy	1 PAGPWGPNGKDGKV 15			
Db	871 PAGPPGPAGKDGPKG 885			

RESULT 19							
062032		PRELIMINARY;	PRT;	1459 AA.			
ID 062032							
AC 062032:							
DT 01-NOV-1996 (TREMBrel. 01, Created)							
DT 01-NOV-1996 (TREMBrel. 01, Last sequence update)							
DT 01-DEC-2001 (TREMBrel. 19, Last annotation update)							
DE PRO-ALPHA-I TYPE II COLLAGEN.							
GN COL2A1.							
OS Mus musculus (Mouse).							
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi;							
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.							
OX NCBI_TAXID=10090;							
RN [1]							
RP SEQUENCE FROM N.A.							
RC STRAIN=<57/BLACK;							
RX MEDLINE=91350489; PubMed=18056113;							
RA Mesaranta M., Toman D., de Crombrugge B., Vuorio E.;							
RT "Mouse type II collagen gene. Complete nucleotide sequence, exon							
RT structure, and alternative splicing.",							
RL J. Biol. Chem. 266:16862-16869(1991).							
RN [2]							
SEQUENCE FROM N.A.							
RC STRAIN=<57/BLACK;							
RA Vuorio E.;							
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.							
DR EMBL; M5161; AAA81011; -.							
DR MGD: MGI:88432; Coll2a1.							
DR InterPro; IPR000077; Collagen.							
DR InterPro; IPR00085; Fib_colagen_C.							
DR InterPro; IPR00107; WFPC.							
DR Pfam; PF01410; COLFI; 1.							
DR Pfam; PF01391; Collagen; 18.							
DR Prodrom; PDD00278; Fib_colagen_C; 1.							
DR SMART; SM00038; COLIFI; 1.							
DR SMART; SM00214; WFPC; 1.							
DR PROSITE; PS01208; WFPC; 1.							
KW Collagen.							
SQ 1459 AA; 139071 MW; A09D24BF7357C827 CRC64;							
Query Match 62.0%; Score 57; DB 11; Length 1459;							
Best Local Similarity 73.3%; Pred. No. 4.4;							
Matches 11; Conservative 0; Mismatches 4; Indels 0; Gaps 0;							
Qy 1 PAGPWGPNGKDGKVG 15							
Db 888 PAGPPGGACKDGPKG 902							
RESULT 20							
093208		PRELIMINARY;	PRT;	303 AA.			
ID 093208							
AC 093208;							
DT 01-FEB-1997 (TREMBrel. 02, Created)							
DT 01-DEC-1997 (TREMBrel. 02, Last sequence update)							
DT 01-DEC-2001 (TREMBrel. 19, Last annotation update)							
DE C11A1.1 PROTEIN.							
GN C11A1.1.							
OS Caenorhabditis elegans.							
OC Eukaryota; Metazoa; Chromadorea; Rhabditida; Rhabditoidea;							
OC Rhabditidae; Peloderrinae; Caenorhabditis.							
OX NCBI_TAXID=6239;							
RN [1]							
RP SEQUENCE FROM N.A.							
RA Gardner A.E.;							
RA Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.							
RN [2]							
RP SEQUENCE FROM N.A.							
RX MEDLINE=99069013; PubMed=9851916;							
RA none;							
RESULT 21							
Q9QBK6		PRELIMINARY;	PRT;	588 AA.			
ID Q9QBK6							
AC Q9QBK6;							
DT 01-MAY-2000 (TREMBrel. 13, Created)							
DT 01-MAY-2000 (TREMBrel. 13, Last sequence update)							
DT 01-DEC-2001 (TREMBrel. 19, Last annotation update)							
DE LAMPENT MEMBRANE PROTEIN 1 (FRAGMENT).							
GN LMP1.							
OS Cynomolgus Epstein-Barr Virus tsB-B6.							
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;							
OC Gammaherpesvirinae; Lymphocryptovirus.							
OX NCBI_TAXID=104240;							
RN [1]							
RP SEQUENCE FROM N.A.							
RX MEDLINE=90250747; PubMed=2160013;							
RA Fujimoto K., Terato K., Miyamoto J., Ishiko H., Fujisaki M., Cho F.,							
RA Honjo S.;							
RT "Establishment of a B-lymphoblastoid cell line infected with Epstein-Barr-related virus from a cynomolgus monkey (Macaca fascicularis).";							
RT J. Med. Primatol. 19:21-30(1990).							
RN [2]							
RP SEQUENCE FROM N.A.							
RA Faucher S., Wright K.E.;							
RA Submitted (ADG-1999) to the EMBL/GenBank/DBJ databases.							
DR EMBL: AF181716; ADD56946; 1. -.							
DR InterPro; IPR02932; EG5SHELL.							
DR InterPro; IPR00188; GABA_receptor.							
DR Pfam; PF02932; NEUR_CHAN_memb; 1.							
DR PRINTS; PRO1228; EGGSHELL.							
FT NON_TER							
SQ SEQUENCE 588 AA; 59338 MW; 62D85879901749BD CRC64;							
Query Match 60.9%; Score 56; DB 12; Length 588;							
Best Local Similarity 60.0%; Pred. No. 2.4;							
Matches 9; Conservative 0; Mismatches 6; Indels 0; Gaps 0;							
Qy 1 PAGPWGPNGKDGKVG 15							
Db 470 PWGPWGPRCPSSNVG 484							
RESULT 22							
Q9796		PRELIMINARY;	PRT;	888 AA.			
ID Q9796							
AC Q9796;							
DT 01-NOV-1996 (TREMBrel. 01, Created)							
DT 01-NOV-1996 (TREMBrel. 01, Last sequence update)							
DT 01-DEC-2001 (TREMBrel. 19, Last annotation update)							
DE ALPHA-1 TYPE XI COLLAGEN (FRAGMENT).							
OX Galius galius (Chicken).							
OC Galius galius.							
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;							
OC Galius.							
OX NCBI_TAXID=9011;							
RN [1]							

RP	SEQUENCE FROM N.A.		OX	NCBI_TAXID=5855;
RC	STRAIN=WHITE LEISHORN; TISSUE=LIMB;		RN	[1]-
RX	MEDLINE=93054557; PubMed=1429507;		RP	SEQUENCE FROM N.A.
RA	Nah H.-D.; Barenbaum M.; Upholt W.B.;		RC	STRAIN=SALVADOR I;
RT	"The chicken alpha 1 (XI) collagen gene is widely expressed in embryonic tissues.";		RX	MEDLINE=96408670; PubMed=8813677;
RL	J. Biol. Chem. 267:22581-22586(1992).		RA	Kaslow D.C., Barr P.J., Crawford K.A., Holey L.J., Landsberg K.E., Gibson H.L., EMBL; M8853; AAA48071; -.
DR	InterPro; IPR000885; Fib_collagen_C.		RT	"Identification and cloning of a locus of serine repeat antigen (sera)-related genes from Plasmodium vivax."
DR	Pfam; PF01410; COLEFI_1.		RL	Mol. Biochem. Parasitol. 78:55-65(1996).
DR	Pfam; PF01391; Collagen_10.		DR	EMBL; U51723; AAB41487.1; -.
DR	ProDom; PD002078; Fib_collagen_C; 1.		DR	InterPro; IPR01064; Crystallin.
DR	SMART; SM00038; COLFI_1.		DR	InterPro; IPR00668; Peptidase_C1.
KW	Collagen.		DR	Pfam; PF00112; Peptidase_C1; 1.
FT	NON_TER 1 888 AA; 86409 MW; B2EC9A0B20F903E CRC64;		DR	PRINTS; PRO0705; PARIN.
SQ	SEQUENCE 888 AA; 308 PRGPOCPNGADGPQG 322		DR	PROSITE; PS00235; CRYSTALLIN_BETAGAMMA; UNKNOWN_1.
RESULT 23			SQ	SEQUENCE 1076 AA; 115597 MW; Q88C2509C3CC17AB_CRC64;
044367			Query Match	Query Match
ID	044367 PRELIMINARY; PRT; 922 AA.		Best Local Similarity	60.9%; Score 56; DB 5; Length 1076;
AC	044367; 01-JUN-1998 (TREMBLrel. 06, Created)		Matches	Best Local Similarity 66.7%; Pred. No. 4.5%; Length 1076;
DT	01-JUN-1998 (TREMBLrel. 06, Last sequence update)		Qy	1 PAGPWPNKGDKGVG 15
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)		Db	843 PDGPAGPSGDGNVG 857
DE	PRECOLLAGEN D.		Query Match	60.9%; Score 56; DB 5; Length 1076;
GN	Mytilus edulis (Blue mussel); Bivalvia; Pteriomorphia; Mytilioida; Eukaryota; Metazoa; Mollusca; Mytilidae; Mytilus.		Best Local Similarity	66.7%; Pred. No. 4.5%; Length 1076;
OS			Matches	1; Mismatches 4; Indels 0; Gaps 0;
OC			Qy	1 PAGPWPNKGDKGVG 15
OC			Db	1096 PPGPVGPGSGKDGNSNG 1110
OX			Query Match	60.9%; Score 56; DB 5; Length 1076;
NCBI_TAXID=6550;			Best Local Similarity	66.7%; Pred. No. 4.5%; Length 1076;
RN	{1}-		Matches	1; Mismatches 4; Indels 0; Gaps 0;
RP	SEQUENCE FROM N.A.		Qy	1 PAGPWPNKGDKGVG 15
RC	TISSUE=FOOT;		Db	843 PDGPAGPSGDGNVG 857
RX	MEDLINE=9807044; PubMed=9405478;		Query Match	60.9%; Score 56; DB 5; Length 1076;
RA	Qin X.X., Coyne K.J., Waite J.H.;		Best Local Similarity	66.7%; Pred. No. 4.5%; Length 1076;
RT	"Tough tendons. Mussel byssus has collagen with silk-like domains.";		Matches	1; Mismatches 4; Indels 0; Gaps 0;
RL	J. Biol. Chem. 272:32623-32627(1997).		Qy	1 PAGPWPNKGDKGVG 15
DR	EMBL; AF029249; AAB86638.1; -.		Db	1096 PPGPVGPGSGKDGNSNG 1110
DR	InterPro; IPR00087; Collagen.		Query Match	60.9%; Score 56; DB 5; Length 1076;
DR	Pfam; PF01391; Collagen; 7.		Best Local Similarity	66.7%; Pred. No. 4.5%; Length 1076;
KW	Collagen.		Matches	1; Mismatches 4; Indels 0; Gaps 0;
SO	SEQUENCE 922 AA; 80306 MW; 599D155E47A2C24A_CRC64;		Qy	1 PAGPWPNKGDKGVG 15
SO	Best Local Similarity 66.7%; Score 56; DB 5; Length 922;		Db	1096 PPGPVGPGSGKDGNSNG 1110
Matches	10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;		Query Match	60.9%; Score 56; DB 5; Length 1076;
Qy	1 PAGPWPNKGDKGVG 15		Best Local Similarity	66.7%; Pred. No. 4.5%; Length 1076;
Db	444 PAGPAGPSGDGNVG 458		Matches	1; Mismatches 4; Indels 0; Gaps 0;
RESULT 24			Qy	1 PAGPWPNKGDKGVG 15
Q26154	PRELIMINARY; PRT; 1076 AA.		Db	1096 PPGPVGPGSGKDGNSNG 1110
ID	026154		Query Match	60.9%; Score 56; DB 5; Length 1076;
AC	026154; 01-NOV-1996 (TREMBLrel. 01, Created)		Best Local Similarity	66.7%; Pred. No. 4.5%; Length 1076;
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)		Matches	1; Mismatches 4; Indels 0; Gaps 0;
DE	V-SERA 3.		Qy	1 PAGPWPNKGDKGVG 15
GN			Db	1096 PPGPVGPGSGKDGNSNG 1110
OS	Plasmodium vivax.		Query Match	60.9%; Score 56; DB 5; Length 1076;
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.		Best Local Similarity	66.7%; Pred. No. 4.5%; Length 1076;
RT			Matches	1; Mismatches 4; Indels 0; Gaps 0;
RESULT 25			Qy	1 PAGPWPNKGDKGVG 15
09W789			Db	843 PDGPAGPSGDGNVG 857
ID	09W789 PRELIMINARY; PRT; 1418 AA.		Query Match	60.9%; Score 56; DB 5; Length 1076;
AC	09W789; 01-NOV-1999 (TREMBLrel. 12, Created)		Best Local Similarity	66.7%; Pred. No. 4.5%; Length 1076;
DT	01-NOV-1999 (TREMBLrel. 12, Last sequence update)		Matches	1; Mismatches 4; Indels 0; Gaps 0;
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)		Qy	1 PAGPWPNKGDKGVG 15
DE	ALPHA TYPE II COLLAGEN.		Db	1096 PPGPVGPGSGKDGNSNG 1110
OS	Cynops pyrrhogaster (Japanese common newt).		Query Match	60.9%; Score 56; DB 5; Length 1076;
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Caudata; Salamandroidea; Salamandridae; Cynops.		Best Local Similarity	66.7%; Pred. No. 4.5%; Length 1076;
OX	NCBI_TAXID=8330;		Matches	1; Mismatches 4; Indels 0; Gaps 0;
RN	{1}-		Qy	1 PAGPWPNKGDKGVG 15
RP	SEQUENCE FROM N.A.		Db	1096 PPGPVGPGSGKDGNSNG 1110
RC	TISSUE=FOOT;		Query Match	60.9%; Score 56; DB 5; Length 1076;
RX	MEDLINE=9807044; PubMed=9405478;		Best Local Similarity	66.7%; Pred. No. 4.5%; Length 1076;
RA	Qin X.X., Coyne K.J., Waite J.H.;		Matches	1; Mismatches 4; Indels 0; Gaps 0;
RT	"Expression of genes of type I and type II collagen in the formation and development of the blastema of regenerating newt limb.";		Qy	1 PAGPWPNKGDKGVG 15
RT	Dev. Dyn. 216:559-71(1999).		Db	1096 PPGPVGPGSGKDGNSNG 1110
RL	EMBL; AB022045; BAR82043.1; -.		Query Match	60.9%; Score 56; DB 5; Length 1076;
DR	InterPro; IPR00087; Collagen.		Best Local Similarity	66.7%; Pred. No. 4.5%; Length 1076;
DR	InterPro; IPR000885; Fib_collagen_C.		Matches	1; Mismatches 4; Indels 0; Gaps 0;
DR	Pfam; PF01410; COLEFI_1.		Qy	1 PAGPWPNKGDKGVG 15
DR	Pfam; PF01391; Collagen; 18.		Db	1096 PPGPVGPGSGKDGNSNG 1110
DR	ProDom; PD002078; Fib_collagen_C; 1.		Query Match	60.9%; Score 56; DB 5; Length 1076;
KW	Collagen.		Best Local Similarity	66.7%; Pred. No. 4.5%; Length 1076;
SO	SEQUENCE 1418 AA; 135067 MW; C19A6E601A2A71E_CRC64;		Matches	1; Mismatches 4; Indels 0; Gaps 0;
SO	Best Local Similarity 66.7%; Score 56; DB 5; Length 1418;		Qy	1 PAGPWPNKGDKGVG 15
Matches	10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;		Db	1096 PPGPVGPGSGKDGNSNG 1110
Qy	1 PAGPWPNKGDKGVG 15		Query Match	60.9%; Score 56; DB 5; Length 1418;
Db	444 PAGPAGPSGDGNVG 458		Best Local Similarity	66.7%; Pred. No. 4.5%; Length 1418;
RESULT 26			Matches	1; Mismatches 4; Indels 0; Gaps 0;
Q90W37			Qy	1 PAGPWPNKGDKGVG 15
ID	Q90W37 PRELIMINARY; PRT; 1420 AA.		Db	1096 PPGPVGPGSGKDGNSNG 1110
AC	Q90W37; 01-DEC-2001 (TREMBLrel. 19, Created)		Query Match	60.9%; Score 56; DB 5; Length 1418;
DT	01-NOV-1996 (TREMBLrel. 01, Last sequence update)		Best Local Similarity	66.7%; Pred. No. 4.5%; Length 1418;
DE	V-SERA 3.		Matches	1; Mismatches 4; Indels 0; Gaps 0;
GN			Qy	1 PAGPWPNKGDKGVG 15
OS	Plasmodium vivax.		Db	1096 PPGPVGPGSGKDGNSNG 1110
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.		Query Match	60.9%; Score 56; DB 5; Length 1418;
RT			Best Local Similarity	66.7%; Pred. No. 4.5%; Length 1418;
RT			Matches	1; Mismatches 4; Indels 0; Gaps 0;
RESULT 27			Qy	1 PAGPWPNKGDKGVG 15
Q90W37			Db	1096 PPGPVGPGSGKDGNSNG 1110
ID	Q90W37 PRELIMINARY; PRT; 1420 AA.		Query Match	60.9%; Score 56; DB 5; Length 1418;
AC	Q90W37; 01-DEC-2001 (TREMBLrel. 19, Created)		Best Local Similarity	66.7%; Pred. No. 4.5%; Length 1418;
DT	01-NOV-1996 (TREMBLrel. 01, Last sequence update)		Matches	1; Mismatches 4; Indels 0; Gaps 0;
DE	V-SERA 3.		Qy	1 PAGPWPNKGDKGVG 15
GN			Db	1096 PPGPVGPGSGKDGNSNG 1110
OS	Plasmodium vivax.		Query Match	60.9%; Score 56; DB 5; Length 1418;
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.		Best Local Similarity	66.7%; Pred. No. 4.5%; Length 1418;
RT			Matches	1; Mismatches 4; Indels 0; Gaps 0;

GN COL2A1.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauvia; Aves; Neognathae; Galliformes; Phasianidae;
 OC Gallus.
 OX NCBI_TAXID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=ETERNAL;
 RA Caixa X., Yongzhi X., Siqi G., Yiyang S.;
 KW SIGNAL; Collagen.
 SEQUENCE 1420 AA: 134999 MW: 88D9AB17F214FF5 CRC64;
 Query Match 60.9%; Score 56; DB 13; Length 1420;
 Best Local Similarity 66.7%; Pred. No. 6; Mismatches 1; Indels 4; Gaps 0;
 Matches 10; Conservative 1; Mismatches 1; Indels 4; Gaps 0;
 QY 1 PAGPWGPNKGDKGVG 15
 Db 1098 PPGPVGPGSKDGGSNG 1112

RESULT 27
 ID 091717 PRELIMINARY; PRT: 1486 AA.
 AC 091717 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DE ALPHA-1 TYPE II COLLAGEN.
 GN COL2A1.
 OS xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Xenopus.
 OC Xenopodinae; Xenopus.
 OX NCBI_TAXID=8355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92011888; PubMed=1918153;
 RA Su M.W., Suzuki H.R., Bleker J.J., Solursh M., Ramirez F.;
 RT "Expression of two nonallelic type II procollagen genes during Xenopus
 laevis embryogenesis is characterized by stage-specific production of
 alternatively spliced transcripts.";
 RL J. Cell Biol. 115:565-575(1991).
 DR EMBL: M63596; AAA49679.1;
 DR InterPro: IPR000087; Collagen.
 DR InterPro: IPR000885; Fib_Collagen_C.
 DR Pfam: PF01391; COLFI; 1.
 DR Pfam: PF00093; vwc; 1.
 DR PRODOM: PD002078; Fib_Collagen_C; 1.
 DR SMART: SM00038; COLFI; 1.
 DR PROSITE: PS01208; WFPC; 1.

Query Match 60.9%; Score 56; DB 13; Length 1491;
 Best Local Similarity 66.7%; Pred. No. 6; Mismatches 1; Indels 4; Gaps 0;
 Matches 10; Conservative 1; Mismatches 1; Indels 4; Gaps 0;
 QY 1 PAGPWGPNKGDKGVG 15
 Db 1170 PPGPVGPGSKDGGSNG 1184

RESULT 29
 ID 091717 PRELIMINARY; PRT: 1767 AA.
 AC 091717 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DE COLLAGEN TYPE XI ALPHA-1.
 GN COL11A1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TAXID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=0455528; PubMed=10486316;
 RA Annunen S., Korikko J., Czarny M., Warman M.L., Brunner H.G.,
 RA Kaarainen H., Mulliken J.B., Tranbjörberg L., Brooks D.G., Cox G.F.,
 RA Crusberg J.R., Curtis M.A., Davenport S.L.H., Friedrich C.A.,
 RA Kaitila I., Kawczynski M.R., Larsen-Bielens A., Mukai S.,
 RA Olsen B.R., Shanno N., Somer M., Viikula M., Zlotogora J.,
 RA Prockop D.J., Ala-Kokko L.;
 RT "Splicing Mutations of 54 bp Exons in the COL11A1 Gene Cause Marshall
 Syndrome but Other Mutations Cause Overlapping Marshall/Stickler
 Phenotypes";
 RT Am. J. Hum. Genet. 65:974-983(1999).
 RL EMBL: AF01112; AAFO4724.1;
 DR EMBL; AF01079; AAFO4724.1; JOINED.
 DR EMBL; AF101080; AAFO4724.1; JOINED.
 DR EMBL; AF101081; AAFO4724.1; JOINED.
 DR EMBL; AF101082; AAFO4724.1; JOINED.
 DR EMBL; AF101083; AAFO4724.1; JOINED.

DR EMBL; AF101084; AAF04724.1; JOINED.
 DR EMBL; AF101085; AAF04724.1; JOINED.
 DR EMBL; AF101086; AAF04724.1; JOINED.
 DR EMBL; AF101087; AAF04724.1; JOINED.
 DR EMBL; AF101088; AAF04724.1; JOINED.
 DR EMBL; AF101089; AAF04724.1; JOINED.
 DR EMBL; AF101090; AAF04724.1; JOINED.
 DR EMBL; AF101091; AAF04724.1; JOINED.
 DR EMBL; AF101092; AAF04724.1; JOINED.
 DR EMBL; AF101093; AAF04724.1; JOINED.
 DR EMBL; AF101094; AAF04724.1; JOINED.
 DR EMBL; AF101095; AAF04724.1; JOINED.
 DR EMBL; AF101096; AAF04724.1; JOINED.
 DR EMBL; AF101097; AAF04724.1; JOINED.
 DR EMBL; AF101098; AAF04724.1; JOINED.
 DR EMBL; AF101099; AAF04724.1; JOINED.
 DR EMBL; AF101100; AAF04724.1; JOINED.
 DR EMBL; AF101101; AAF04724.1; JOINED.
 DR EMBL; AF101102; AAF04724.1; JOINED.
 DR EMBL; AF101103; AAF04724.1; JOINED.
 DR EMBL; AF101104; AAF04724.1; JOINED.
 DR EMBL; AF101105; AAF04724.1; JOINED.
 DR EMBL; AF101106; AAF04724.1; JOINED.
 DR EMBL; AF101107; AAF04724.1; JOINED.
 DR EMBL; AF101108; AAF04724.1; JOINED.
 DR EMBL; AF101109; AAF04724.1; JOINED.
 DR EMBL; AF101110; AAF04724.1; JOINED.
 DR EMBL; AF101111; AAF04724.1; JOINED.
 DR InterPro; IPR000087; Collagen.
 DR InterPro; IPR000885; Fib.collagen_C.
 DR InterPro; IPR001701; laminin_G.
 DR InterPro; IPR003129; TSPN.
 DR Pfam; PF01410; COIFI; 1.
 DR Pfam; PF02210; TSPN; 1.
 DR Pfam; PF00210; TSPN; 1.
 DR SMART; SM00038; COIFI; 1.
 DR SMART; SM00282; LammG; 1.
 DR SMART; SM00210; TSPN; 1.
 SQ SEQUENCE 1767 AA; 176644 MW; 72E00F4CAA731B26 CRC64;

Query Match 60.9%; Score 56; DB 4; Length 1767;
 Best Local Similarity 66.7%; Pred No. 7.6; DB 10; Conservative
 Matches 10; Mismatches 0; Indels 5; Gaps 0; Gaps 0;

QY 1 PAGPWGPNGKDKGVG 15
 DB 1187 PRGPQGPNGADGPGQG 1201

RESULT 30

Q9UUT5 PRELIMINARY; PRT; 1806 AA.
 ID Q9UUT5 AC

DT 01-MAY-2000 (TREMBL; 13, Last sequence update)
 DT 01-DEC-2001 (TREMBL; 19, last annotation update)

DE COLLAGEN TYPE XI ALPHA-1 ISOFORM A.

GN COL1A1.

OS Homo sapiens (Human).

OC Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; NCBI_TAXID=9606; RN [1]

SEQUENCE FROM N.A.

MEDLINE=0455728; PubMed=10486316;

Annunen S., Korkko J., Czarny M., Warnman M.L., Brunner H.G., Cruysberg J.R., Curtis M.A., Davenport S.L.H., Friedrich C.A.,

Kaitila I., Krawczynski M.R., Latos-Bielecka A., Mukai S., Olsen B.R., Shino N., Somer M., Vilkku M., Zlotogora J., Prockop D.J., Ala-Kokko L.,

"Splicing Mutations of 54 bp Exons in the COL1A1 Gene Cause Marshall

RT Syndrome but Other Mutations Cause Overlapping Marshall/Stickler Phenotypes.", Ann. J. Hum. Genet. 65:974-983 (1999).

RT PhenoTypes.; DR

RL Ann. J. Hum. Genet. 65:974-983 (1999).

DR EMBL; AF101112; AAF04725.1; JOINED.

DR EMBL; AF101079; AAF04725.1; JOINED.

DR EMBL; AF101080; AAF04725.1; JOINED.

DR EMBL; AF101031; AAF04725.1; JOINED.

DR EMBL; AF101082; AAF04725.1; JOINED.

DR EMBL; AF101083; AAF04725.1; JOINED.

DR EMBL; AF101084; AAF04725.1; JOINED.

DR EMBL; AF101085; AAF04725.1; JOINED.

DR EMBL; AF101086; AAF04725.1; JOINED.

DR EMBL; AF101087; AAF04725.1; JOINED.

DR EMBL; AF101088; AAF04725.1; JOINED.

DR EMBL; AF101089; AAF04725.1; JOINED.

DR EMBL; AF101090; AAF04725.1; JOINED.

DR EMBL; AF101091; AAF04725.1; JOINED.

DR EMBL; AF101092; AAF04725.1; JOINED.

DR EMBL; AF101093; AAF04725.1; JOINED.

DR EMBL; AF101094; AAF04725.1; JOINED.

DR EMBL; AF101095; AAF04725.1; JOINED.

DR EMBL; AF101096; AAF04725.1; JOINED.

DR EMBL; AF101097; AAF04725.1; JOINED.

DR EMBL; AF101098; AAF04725.1; JOINED.

DR EMBL; AF101099; AAF04725.1; JOINED.

DR EMBL; AF101100; AAF04725.1; JOINED.

DR EMBL; AF101101; AAF04725.1; JOINED.

DR EMBL; AF101102; AAF04725.1; JOINED.

DR EMBL; AF101103; AAF04725.1; JOINED.

DR EMBL; AF101104; AAF04725.1; JOINED.

DR EMBL; AF101105; AAF04725.1; JOINED.

DR EMBL; AF101106; AAF04725.1; JOINED.

DR EMBL; AF101107; AAF04725.1; JOINED.

DR EMBL; AF101108; AAF04725.1; JOINED.

DR EMBL; AF101109; AAF04725.1; JOINED.

DR EMBL; AF101110; AAF04725.1; JOINED.

DR EMBL; AF101111; AAF04725.1; JOINED.

DR EMBL; AF00210; TSPN; 1.

DR Pfam; PF00210; TSPN; 1.

DR SMART; SM00038; COIFI; 1.

DR SMART; SM00282; LammG; 1.

DR SMART; SM00210; TSPN; 1.

DR Pfam; PF01391; Collagen; 18.

DR InterPro; IPR000087; Collagen.

DR InterPro; IPR001791; Laminin_G.

DR InterPro; IPR003129; TSPN.

DR Pfam; PF01410; COIFI; 1.

DR SMART; SM00038; COIFI; 1.

DR SMART; SM00282; LammG; 1.

DR SMART; SM00210; TSPN; 1.

Query Match 60.9%; Score 56; DB 4; Length 1806;
 Best Local Similarity 66.7%; Pred No. 7.7; DB 10; Conservative
 Matches 10; Mismatches 0; Indels 5; Gaps 0; Gaps 0;

QY 1 PAGPWGPNGKDKGVG 15
 DB 1226 PRGPQGPNGADGPGQG 1240

Search completed: November 1, 2002, 12:53:47

Job time : 24.5 secs